PROBLEMS OF MODERN PHYSICS

A COURSE OF LECTURES DELIVERED IN THE CALIFORNIA INSTITUTE OF TECHNOLOGY

BY

H. A. LORENTZ

PROFESSOR IN THE UNIVERSITY OF LEIDEN

EDITED BY

H. BATEMAN

PROFESSOR IN THE CALIFORNIA INSTITUTE OF TECHNOLOGY

GINN AND COMPANY

BOSTON · NEW YORK · CHICAGO · LONDON
ATLANTA · DALLAS · COLUMBUS · SAN FRANCISCO

PRINTED IN THE UNITED STATES OF AMERICA

427.2

PREFACE

The publication of these lectures, which I delivered at the California Institute of Technology in the beginning of 1922, and which Professor H. Bateman kindly prepared for the press shortly afterwards, has been unduly delayed. Several important questions which I should have liked to treat more fully had been only slightly touched upon, and I therefore wished to complete the work by the addition of a certain number of notes, such as will be found in the Appendix. To my regret the preparation of these fell in a period during which my time was much taken by other work that could not be deferred.

Even as the book is now it cannot be said to be in any way complete. Indeed, of late years the progress of physics has been so rapid, and so many new ideas, especially about the structure of atoms and their radiation, have sprung up, that the development that has taken place since 1922 might in itself form the subject for a treatise. I hope, however, that my lectures may be of some use as an introduction to the problems on which the efforts of physicists are now largely concentrated.

My hearty thanks are due to Professor Bateman for the valuable help he has given me and to the publishers for the great care they have bestowed upon the work.

H. A. LORENTZ

CONTENTS

PROBLEMS OF MODERN PHYSICS

LIGHT AND MATTER, WITH SOME CONSIDERATIONS ON RELATIVITY

1. The Propagation of Light. In the early lectures I shall have to present to you nothing but quite old-fashioned, or perhaps I may say classical, questions. My subject will be the propagation of light in different media. It will lead us to some interesting problems, the solution of which will be useful to us on later occasions.

I need not speak to you of the older forms of the undulatory theory of light, like those developed by Fresnel, Cauchy, Green, F. E. Neumann, and others, which were current before the time of Maxwell and which depended on different conceptions of an elastic ether. Nowadays we are concerned only with the electromagnetic theory of light, in which there is no longer any discussion of a density or elasticity of the ether. In the electromagnetic theory of light attention is fixed on the electric and magnetic fields that can exist in the "ether." With your permission I shall still use this time-honored word, although nowadays some physicists prefer not to speak of an ether. At all events we must be careful not to assign to this medium so much of the properties of ordinary matter as was done in old times. Great though the change has been in the physical point of view, the mathematical form in which the theory of light can be presented has remained much the same as it was. This will be seen from the following analysis.

2. Plane Waves of Light. For the sake of simplicity we consider a propagation in one direction only and fix our attention on just one of the quantities used to specify the state of

the ether; for example, one component ϕ of the electric or magnetic force. If, moreover, we assume that, at any instant of time t, ϕ has the same value at all points of a plane, $x =$ constant, perpendicular to the direction of propagation, the manner of propagation is determined by a partial differential equation of the form

$$\frac{\partial^2\phi}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}. \tag{1}$$

This simple equation is exactly similar to the equation to which one is led in the theory of the motion of a stretched string or of the vibrations of sound in a mass of gas. This celebrated equation was solved by D'Alembert in 1747, the general solution being

$$\phi = F(x - ct) + G(x + ct), \tag{2}$$

where F and G are arbitrary functions of their arguments $x - ct$ and $x + ct$. A particular solution is obtained by taking only one of the terms; thus,

$$\phi = F(x - ct) \tag{3}$$

corresponds to a propagation with velocity c in the direction of the positive x. This may be seen by remarking that $x - ct$ and ϕ remain unchanged when t and x are increased by the quantities Δt and $c\Delta t$ respectively. Thus, if the state of the ether corresponding to a definite value of ϕ is found at time t at some point P, that same value will be found at the later time $t + \Delta t$ at a point Q such that PQ is in the direction of the positive x and has the length $c\Delta t$.

This is the purest type of propagation that can be imagined, since it takes place without any change in the intensity or form of the wave. This means, in particular, that the velocity of propagation of a periodic train of waves is independent of the period or wave-length, and that a group of waves leaves no residue, or tail, behind it as it passes by.

It is also clear that

$$\phi = G(x + ct) \tag{4}$$

represents a propagation with the same velocity c in the direction of the negative x. The complete expression (2) corresponds to a superposition of the special states just now considered,—

a superposition that is possible because the partial differential equation is linear and homogeneous.

We may also remark that a particular form of (3) is

$$\phi = a \cos n\left(t - \frac{x}{c} + p\right), \tag{5}$$

the well-known expression for the propagation of simply harmonic vibrations; a is the amplitude and n the frequency,* while the constant p determines the phase.

As you well know, the complex solution

$$\phi = a\, e^{in\left(t - \frac{x}{c} + p\right)}, \tag{6}$$

in which $i = \sqrt{-1}$ is the imaginary unit, is often used instead of (5), and we shall sometimes have occasion to do so. When a complex expression of this kind is used, it is implied that the solution of the physical problem may be obtained by taking only the real part of the complex expression. Indeed, it is easily seen that if a complex expression is a solution of a linear homogeneous equation (with real coefficients), the real part of that expression will be so likewise.

3. The Electromagnetic Equations. Thus far I have spoken of only one, ϕ, of the quantities determining the state of the ether. I need scarcely remind you that in reality the state is determined by two vectors, — the electric force, which we shall represent by **E**, and the magnetic force, for which we shall write **H**, the components of these vectors being represented by E_x, E_y, E_z and H_x, H_y, H_z respectively. The relations between these two vectors are expressed by the familiar equations generally known as Maxwell's equations. In vector form these equations are

$$\left.\begin{aligned} \operatorname{div} \mathbf{E} = 0, \quad \operatorname{curl} \mathbf{H} &= \frac{1}{c}\dot{\mathbf{E}} \\ \operatorname{div} \mathbf{H} = 0, \quad \operatorname{curl} \mathbf{E} &= -\frac{1}{c}\dot{\mathbf{H}} \end{aligned}\right\}. \tag{7}$$

* The word "frequency" is used here in the same sense as in Lorentz's "Theory of Electrons," p. 6. The quantity $n/2\pi$ is now generally called the frequency and n the rapidity or speed. See Lamb's "Dynamical Theory of Sound," p. 10.

In terms of the components,

$$\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}=0$$

$$\cdots\cdots\cdots$$

$$\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}=\frac{1}{c}\frac{\partial E_x}{\partial t}$$

$$\cdots\cdots\cdots,$$

the other equations being obtained from those written down by writing **H** for **E** and $-$ **E** for **H** and by making a cyclic permutation of the coördinates and suffixes x, y, z.

The corresponding equations for the interior of a non-conducting ponderable body may be written down also. In this case there are four distinct vectors **E**, **D**, **B**, and **H** (in the well-known notation) connected by the equations

$$\operatorname{div}\mathbf{D}=0, \tag{8}$$

$$\operatorname{curl}\mathbf{H}=\frac{1}{c}\dot{\mathbf{D}}, \tag{9}$$

$$\operatorname{div}\mathbf{B}=0, \tag{10}$$

$$\operatorname{curl}\mathbf{E}=-\frac{1}{c}\dot{\mathbf{B}}. \tag{11}$$

These equations are quite general, being the same for all bodies.

Substances differ from one another electromagnetically in the form of the supplementary equations connecting these vectors which must be added to complete the scheme.

In the case of an isotropic body, which is stationary relative to the axes of coördinates, the supplementary equations or constitutive relations are

$$\mathbf{D}=\epsilon\mathbf{E},\quad \mathbf{B}=\mu\mathbf{H}, \tag{12}$$

where the second equation is an approximation suitable for weak magnetic fields. In non-magnetic bodies μ differs so slightly from unity that the difference between **B** and **H** may be ignored.

The coefficients μ and ϵ are called the *permeability* and *specific inductive capacity* of the medium. The term *dielectric constant* is sometimes used for ϵ, but when we are dealing with the

propagation of light the value of ϵ depends on the frequency (or wave-length) of the light, and so the use of the word *constant* is perhaps undesirable. This variation of ϵ with the wave-length becomes of importance in a theory of dispersion, for, as will be seen presently, the velocity of propagation of light depends on the value of ϵ.

In writing down the above formulæ I must call your attention to the great and wonderful simplification which electrical theory has undergone in the course of the last half century. Formerly electrostatics, magnetism, and electrodynamics were separate subjects, with but loose connections between them, and in the last-named there were many different theories, like those of Wilhelm Weber, Gauss and Clausius, Ampère and Grassmann. Now all has been blended into one theory, the main equations of which can be written on a page of a pocket notebook. That we have got so far is due in the first place to Maxwell, and next to him to Heaviside and Hertz.

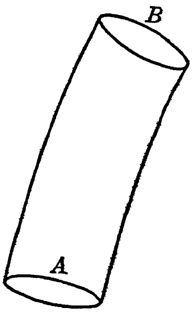

FIG. 1

The meaning of the equations will not require long explanation. Equation (9) expresses the relation between the electric current and the magnetic field, and in (11) we have the general law for induced currents. Equations (8) and (10) show that both the dielectric displacement **D** and the magnetic induction **B** are solenoidally distributed. This means that both vectors are analogous to the velocity of an incompressible fluid. If we consider a tube of flow in such a fluid, the inward flow of fluid (Fig. 1) across a section *A* is just balanced by the outward flow across the section *B*, there being by definition no flow across the sides of the tube. The analogous property of tubes of dielectric displacement and of induction may be regarded as the physical interpretation of the mathematical equation.

The vector **E** may be defined as the force on a unit electric charge, and **H** as the force on a unit magnetic pole; but these definitions need amplification, as it is necessary to specify in each case the type of cavity surrounding the point at which

the force is estimated. Indeed, by suitably changing the form of the cavity we can obtain definitions of **D** and **B**. We need not enter into details here, as these matters are discussed in textbooks on electricity and magnetism.

The vector **D** may be considered as a displacement of electricity from its position of equilibrium, measured by the amount that has passed through unit of area, in exactly the same sense as a current is held to consist in a flow of electricity. In ponderable bodies **D** is composed of two parts, the first of which is the electric polarization **P** of the matter, whereas the second part, which has its seat in the ether, is equal in direction and magnitude (on account of the choice of rational units) to the electric force **E**; that is

$$\mathbf{D} = \mathbf{E} + \mathbf{P}. \tag{13}$$

This equation is exactly similar to the well-known relation

$$\mathbf{B} = \mathbf{H} + \mathbf{M} \tag{14}$$

between the magnetic force **H**, the magnetic induction **B**, and the magnetization **M**.

To some extent the vectors are defined by the equations which they satisfy and the relations between **E**, **D**, **B**, and **H**.

The simple relations given above are not suitable for all types of substance. In a crystal the relation between **D** and **E** may, with a suitable choice of coördinate axes, be written in the form

$$D_x = \epsilon_1 E_x, \quad D_y = \epsilon_2 E_y, \quad D_z = \epsilon_3 E_z, \tag{15}$$

where the coefficients ϵ_1, ϵ_2, ϵ_3 are generally different.

The inequality of these coefficients leads to double refraction. If we wish to develop a theory of this phenomenon, we seek first a solution that represents a propagation of plane waves.

4. Plane Waves of Light in a Crystal. Putting

$$(E_x, E_y, E_z) = (f, g, h)\, e^{in\left(t - \frac{x\cos\alpha + y\cos\beta + z\cos\gamma}{v} + p\right)} \tag{16}$$

and substituting in the differential equations, we obtain expressions * for the components of the other vectors, and it is then

* See Note 1, Appendix.

found that there are certain relations between the constants $f, g, h, \alpha, \beta, \gamma, v$. For given values of α, β, γ, that is, for a given direction of the normal to the waves, we can calculate the velocity v and the ratios of f, g, and h.

It is found that for any given direction of propagation there are generally two values of v and corresponding values of the ratios $f : g : h$.

If the coefficients ϵ were really constants, there would be no dispersion. We can arrive at a dispersion of light in the crystal by introducing somewhat more complicated relations between **E** and **D**, amounting to this,— that the ϵ's are functions of the frequency n.

As before, the state of the ether is the same at all points of a plane perpendicular to the direction of propagation, and so the waves may be called plane waves.

5. The Principle of Huygens. In all textbooks of physics we can find an exposition of what is called Huygens's principle, and of the way in which it can be used for determining the propagation of waves. Let us suppose, for instance, that a disturbance of equilibrium that has started from a luminous point at some previous instant has reached the points of the surface σ at the time t. Then, according to the principle, every point of σ will become the center of a disturbance that spreads out in all directions. Hence, if we want to know the state of things at the time $t + dt$, we can reason as follows:

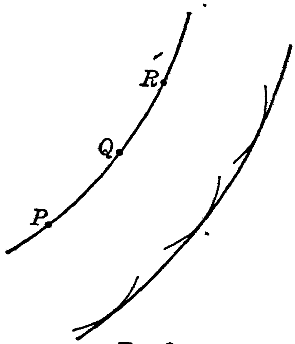

FIG. 2

Around each point, P, Q, R, etc., of σ (Fig. 2) an "elementary wave" has been formed, and at the time $t + dt$ the disturbance will therefore have reached the surface σ' that is tangent to all these waves and by which they are consequently enveloped.

You will not fail to observe that in this form the principle is somewhat vague. If our statement of it is to be made quite

definite, we must be able to describe much more accurately the motion in the elementary waves and the way in which it depends on the disturbance that exists in the surface σ. Moreover, the principle has to be proved. All this has been done by Kirchhoff* and other physicists. We shall perhaps return to these questions later on; for the moment a somewhat superficial treatment will suffice.

6. Applications of Huygens's Principle. The wide applicability of the principle may be seen from the following examples:

1. *Isotropic Media.* The elementary waves are in this case spheres. In Fig. 2 the spheres have their centers at the points P, Q, R, $\cdots$, and all have the same radius cdt; the surface σ' may then be said to be parallel to the surface σ.

2. *Crystals.* The elementary waves have the well-known less simple form of Fresnel's wave-surface; in the case of a uniaxial crystal this consists of a sphere and an ellipsoid.

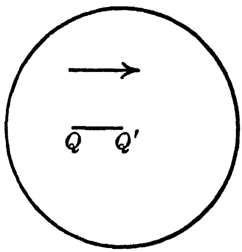

FIG. 3

3. *Medium in Motion (Theory of Aberration).* If the medium is isotropic, the difference between this case and 1 is that the elementary wave formed around a point Q, though still a sphere, will no longer have its center at Q. If we suppose the medium to be moving through our diagram, the elementary wave will be dragged along with it to a greater or less extent, according to the value of Fresnel's dragging coefficient. The center will thus be shifted to Q' (Fig. 3).

4. *Diffraction.* Let us suppose that the surface σ is not fully illuminated but only partly illuminated, as, for instance, in the case of a screen with a hole in it. We then have the phenomenon of *diffraction,* which was first adequately treated, with the aid of Huygens's principle, by Fresnel. This provides, indeed, a most important application of the principle, which will be discussed in some detail.

Let a beam of light fall on a screen with an opening AB (Fig. 4). The points lying in the plane of the screen within AB

***Ann. d. Phys. u. Chem.*, Bd. 18 (1883), p. 663; Kirchhoff's proof is described in Lorentz's "Theory of Electrons," Note 4. Another proof, due to Beltrami, is given in Love's "Mathematical Theory of Elasticity," § 210.

become new centers of vibration. If we want the distribution of light in the plane U behind the screen, we have to observe that each point P of U receives light from all points of AB. All the vibrations received by P are superposed, and there is interference depending on the differences of phase determined by the various distances of P from the points of the opening. Thus interference fringes, whose form depends on the shape of the opening, will appear on the screen U.

Huygens's principle implies a tendency of light to spread in all directions, and this is the chief characteristic of the phenomenon of diffraction. If the opening AB is extremely small, there will be but little difference in the intensity with which the vibrations are propagated in different directions. If the opening becomes wider, the propagation begins to be somewhat concentrated in the direction of the incident beam, but the boundary of the beam behind the opening is not sharp, being made more or less indefinite by diffraction. It is only when the dimensions of the opening AB are very large in comparison with the wave-length that, at distances behind the screen which are not too great, the diffraction can be neglected. Then we can speak of rays of light, and (in simple cases) of propagation along straight lines. The reason why observation makes us familiar with this rectilinear propagation is that in everyday experience the dimensions of openings or of bodies that cast shadows are extremely large in comparison with the wave-length. Still, even when we use large openings, the effects of diffraction will become sensible at great distances.

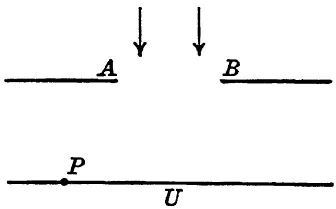

FIG. 4

Let us suppose that an opening of 1 centimeter is taken and that we succeed in making a beam of absolutely parallel rays fall on it. Then at a distance of 1 kilometer behind the beam the distance between the diffraction fringes will be of the order of 6 centimeters, so that there can be no question of a sharply delimited beam. If the opening had a diameter of

10 centimeters the diffraction fringes would have a width of about 0.6 centimeter; the beam would then have rather sharp borders at a distance of 1 kilometer but would no longer be sharp at a distance of 10 kilometers.

This tendency of light to spread in all directions is one of the great difficulties in modern physics. Many phenomena — photo-electricity, for instance — could be easily explained if the energy of light were concentrated in "quanta" each confined to a small space. But, according to the old theory, such a thing cannot be. Energy tends to become more and more dilute. Returning to Huygens's construction, we remark that if the vibrations exist only in a part AB of the wave-front σ, the next wave-front, σ', will be illuminated only so far as it envelops the elementary waves formed around the points within the opening (Fig. 5). The boundary of σ' is determined by the points of contact A', B', $\cdots$ with the extreme elementary waves; namely, those points corresponding to the points on the boundary of AB. Parts of the elementary waves lying beyond the boundary just mentioned may be neglected.

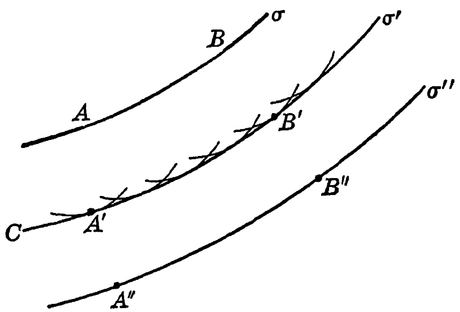

Fig. 5

Proceeding in this way from σ' onward, we obtain new wave-fronts, σ'', σ''', etc., and new points, A'', B'', etc. The line passing through A, A', A'', $\cdots$ is called *a ray of light*. A ray may be defined as the generating line of a tubular surface (in simple cases a cylindrical or conical surface) by which a beam of light can be laterally limited.

Example. If we know the way in which the velocity of waves of light depends on the gravitational potential or potentials (for example, the law by which this velocity changes in the sun's gravitational field, — let us say Einstein's law), then we can find, by Huygens's construction, the bending of the rays of light in that field.

7. Analytical Treatment of Beams of Light. The theory of lateral delimitation, that is, a theory of rays, may be developed independently of Huygens's construction.

It is easily seen that in the case of the ether the fundamental equations (7) admit the solution (the simplest of all)

$$E_y = a \cos n\left(t - \frac{x}{c}\right), \quad H_z = a \cos n\left(t - \frac{x}{c}\right), \tag{17}$$

the values of E_x, E_z, H_x, H_y being zero and a being a constant. These formulæ represent a propagation of plane-polarized light in the direction of x, the magnetic vibrations being at right angles to the electric ones and the system of waves extending indefinitely in a lateral as well as in a longitudinal direction.

Now we can easily find another solution by supposing a to be a function of the transverse coördinates y and z; that is, a function that is *slowly variable.** By this I mean, for instance, that the value of $\frac{\partial E_y}{\partial y}$ that is due to the change of a in the direction of y is much smaller than the value of $\frac{\partial E_y}{\partial x}$ that is due to the change of phase in the direction x. In symbols,

$$\frac{\partial a}{\partial y} << \frac{n}{c} a, \quad \text{or} \quad \frac{\partial a}{\partial y} << \frac{a}{\lambda}, \tag{18}$$

where

$$\lambda = \frac{2\pi c}{n}. \tag{19}$$

This can also be expressed by saying that when we go forward in the wave-front over a distance equal to the wave-length λ, the change in a is very small in comparison with a itself.

We suppose the same for $\frac{\partial a}{\partial z}$; we also suppose that these small quantities are slowly variable in their turn and that the second derivatives $\frac{\partial^2 a}{\partial y^2}$, $\frac{\partial^2 a}{\partial y\, \partial z}$, $\cdots$ are small quantities of the second order, such quantities and quantities of higher orders being neglected, likewise squares and products and higher powers of $\frac{\partial a}{\partial y}$ and $\frac{\partial a}{\partial z}$.

* H. A. Lorentz, Abhandlungen über theoretische Physik, Bd. I, p. 415.

In the new solution E_y and H_z have the values given above, but in addition *

$$\left.\begin{aligned} E_x &= \frac{c}{n}\frac{\partial a}{\partial y}\sin n\left(t-\frac{x}{c}\right), \quad E_z = 0 \\ H_x &= \frac{c}{n}\frac{\partial a}{\partial z}\sin n\left(t-\frac{x}{c}\right), \quad H_y = 0 \end{aligned}\right\}. \tag{20}$$

The way in which a changes in the wave-front from point to point is left indeterminate except for the above restrictions. Since, then, a is an arbitrary function of y and z, it can be zero outside some closed curve. We have therefore a limited beam, which is just what is wanted. Our assumption of slowly varying quantities implies a boundary that is not sharp. This is in the nature of things.

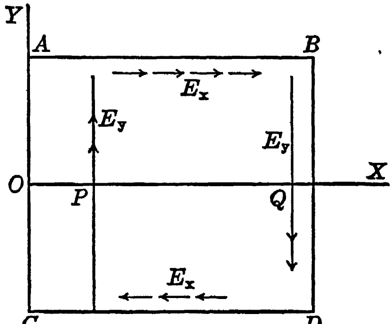

FIG. 6

It is noteworthy that we now have not only an E_y but also an E_x, so that it may at first be thought that this is a violation of the Young-Fresnel principle of transverse vibrations. The true statement of this principle, however, is that the vectors **E** and **H** are solenoidally distributed. This condition is satisfied by our solution, from which we easily derive div $\mathbf{E} = 0$ and div $\mathbf{H} = 0$. In order to illustrate what we have found, we can draw the lines of electric and magnetic force. These are closed curves, somewhat of the kind shown (for the electric force) in Fig. 6, where AB and CD lie in the boundary of the beam. It is supposed that at the instant considered $\cos n\left(t-\frac{x}{c}\right)$ has the value $+1$ at the point P and the value -1 at the point Q, the distance PQ being half the wave-length.

The components E_x and H_x occur only where $\frac{\partial a}{\partial y}$ and $\frac{\partial a}{\partial z}$ are different from zero. They can be confined, as shown in the diagram, to a thin layer near the boundary of the beam, the

* See Note 2, Appendix.

value of a being constant and the vibrations being purely transverse through the greater part of the section.

It should be remarked that along any line normal to the wave-front the amplitude a is constant.

The same mode of reasoning can be applied to doubly refracting media. Then we find that the amplitude a is constant, not as it was just now, along the normal to the wave-front, but along a line having just the direction that we find for a ray of light by Huygens's construction.

We might also apply it to the case of a medium whose properties are not the same at all points; for instance, to the atmosphere or to a solution of salt. In such a case ϵ is not the same at all places, and we cannot have waves in which the amplitudes of both **E** and **D** are the same at all points. Instead the amplitudes vary slowly on account of the variation of the properties of the medium.

Physical media generally have slowly varying properties in the sense that the properties are nearly the same at two points whose distance apart is a wave-length. Light passes through such a medium with no appreciable reflection. To see this let us consider the case of a stratified atmosphere in which the refractive index μ varies from 1 to 1.00029. Let us suppose that the transition takes place in a large number of small steps; then at each step the surface of separation reflects. By Fresnel's well-known formula the ratio of the intensity of the reflected and of the incident light is given, in the case of normal incidence, by

$$\left(\frac{\mu' - \mu''}{\mu' + \mu''}\right)^2,$$

where μ' and μ'' are the values of the index of refraction on both sides of the boundary. We may put

$$\mu' - \mu'' = \frac{\mu - 1}{N} \tag{21}$$

for each of the N surfaces of separation and replace $\mu' + \mu''$ by 2. The portion of the incident light that is reflected by one surface will therefore be

$$\frac{1}{4N^2}(\mu - 1)^2, \tag{22}$$

and we may write for the portion that is reflected by the N surfaces together

$$\frac{1}{4N}(\mu - 1)^2 = \frac{1}{4N}(0.00029)^2, \tag{23}$$

which becomes excessively small when N is large. There is thus no appreciable reflection. Reflection can, in fact, take place only when there is a sharp boundary; that is, when the properties of the medium change appreciably in a distance equal to the wave-length.

In the above calculation it has been supposed that the thickness of the individual layers is very great in comparison with the wave-lengths. Then there are no phenomena of interference, and we have simply to add to each other the reflected intensities. This amounts to a multiplication by N, because it is assumed (and this is confirmed by the result) that the light reaching one of the lower layers is not sensibly weaker than the original light.

8. Group-Velocity. We now pass on to the consideration of group-velocity, which admits of similar mathematical treatment.

We have to consider the case of a dispersive medium and must therefore start from equations that involve an influence of the frequency on the velocity of propagation. These equations may be of a somewhat complicated form, but at all events they will be linear and homogeneous, and it will be possible to eliminate from them all dependent variables but one, say ϕ. The final equation will determine ϕ as a function of t and the three coördinates x, y, z, or as a function of t and x if we suppose ϕ to be independent of y and z.

We shall write down the equation in the symbolical form

$$F(D_t, D_x)\phi = 0, \tag{24}$$

in which D_x and D_t are symbols for the operations $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial t}$, and where $F(D_t, D_x)$ indicates what we may call a polynomial in D_t and D_x with constant coefficients. The polynomial is not

necessarily homogeneous. The meaning of our equation will become clear if we give one or two examples; for instance

$$(pD_t^2+qD_x^2)\phi \quad \text{stands for} \quad p\frac{\partial^2\phi}{\partial t^2}+q\frac{\partial^2\phi}{\partial x^2} \tag{25}$$

and $$(\alpha D_t^3+\beta D_t^2 D_x+\gamma D_t D_x+\delta D_x^2)\phi$$

for $$\alpha\frac{\partial^3\phi}{\partial t^3}+\beta\frac{\partial^3\phi}{\partial t^2\,\partial x}+\gamma\frac{\partial^2\phi}{\partial t\,\partial x}+\delta\frac{\partial^2\phi}{\partial x^2}. \tag{26}$$

It must be noticed that the exponents attached to the symbols D_t and D_x always mean that the operations indicated by them have to be repeated a certain number of times, and that D_tD_x means that the two operations D_x and D_t are to be applied successively.

We assume that equation (24) is linear and homogeneous, so that the principle of superposition may hold. This means in particular that if we have one solution of the equation, we can obtain another by multiplying it by a constant factor. This implies that the course of beams of light and their velocity of propagation are independent of the intensity. Attempts have often been made to discover a variation of the velocity of light with its intensity, but no variation has yet been found; consequently equation (24) must be linear and homogeneous.

Now put $$\phi=a\,e^{in\left(t-\frac{x}{v}\right)} \tag{27}$$

and first suppose that a is a constant; then

$$D_t=\frac{\partial}{\partial t}=in, \quad D_x=\frac{\partial}{\partial x}=-\frac{in}{v}. \tag{28}$$

This means that the differentiation of ϕ (and similarly of its derivatives) with respect to t is equivalent to a multiplication by in, and a differentiation with respect to x amounts to a multiplication by $-\frac{in}{v}$.

The expression (25) now becomes

$$\left[p(in)^2+q\left(-\frac{in}{v}\right)^2\right]\phi,$$

and in general $F(D_t, D_x)$ now becomes the product of ϕ and a function (in the ordinary sense of the word) of in and $-\frac{in}{v}$. The differential equation (24) reduces to

$$F\left(in, -\frac{in}{v}\right) = 0, \tag{29}$$

and this is an algebraic equation by which the value of v for any given frequency will be determined.

We shall next suppose that a is no longer a constant but depends on x and t, being a slowly varying function. This means that the amplitude is not constant all along the beam but shows gradual changes; these may go so far that the amplitude is different from zero only between two wave-fronts at a certain distance from each other, so that we shall really find the way in which a limited group of waves is propagated. The group-velocity w may be regarded as the velocity of the boundary, although strictly the boundary is not sharp. We shall rather determine the velocity with which a definite amplitude a is propagated, and shall define this as the velocity of the group. Instead of (28) we now have

$$\frac{\partial \phi}{\partial t} = in\, a\, e^{in\left(t-\frac{x}{v}\right)} + \frac{\partial a}{\partial t} e^{in\left(t-\frac{x}{v}\right)} = \left(in + \frac{\partial \log a}{\partial t}\right)\phi,$$

that is,

$$\frac{\partial}{\partial t} = in + \frac{\partial \log a}{\partial t}; \tag{30}$$

and similarly

$$\frac{\partial}{\partial x} = -\frac{in}{v} + \frac{\partial \log a}{\partial x}. \tag{31}$$

Substituting in the differential equation, we now get

$$F\left(in + \frac{\partial \log a}{\partial t}, -\frac{in}{v} + \frac{\partial \log a}{\partial x}\right) = 0, \tag{32}$$

or, taking into account that a is slowly variable, so that $\frac{\partial \log a}{\partial t}$ and $\frac{\partial \log a}{\partial x}$ are small in comparison with n and $\frac{n}{v}$,

$$F\left(in, -\frac{in}{v}\right) + F_1 \frac{\partial \log a}{\partial t} + F_2 \frac{\partial \log a}{\partial x} = 0, \tag{33}$$

where the derivatives of $F\left(in, -\frac{in}{v}\right)$ with respect to the arguments in and $-\frac{in}{v}$ are denoted by F_1 and F_2 respectively.

This equation can be satisfied on the assumption that v has the value given by (29), whereas a is determined by

$$F_1 \frac{\partial \log a}{\partial t} + F_2 \frac{\partial \log a}{\partial x} = 0, \tag{34}$$

or

$$F_1 \frac{\partial a}{\partial t} + F_2 \frac{\partial a}{\partial x} = 0. \tag{35}$$

This equation admits of a simple interpretation. At a definite instant the amplitude a may be any function of x, but the value which it has at any later instant is determined by the rule that the value of a is propagated in the direction of x with the velocity

$$w = \frac{F_2}{F_1}. \tag{36}$$

Indeed, if t is made to increase by the infinitely small amount Δt, and x by the amount $w\Delta t$, the change in a will be

$$\left(\frac{\partial a}{\partial t} + w \frac{\partial a}{\partial x}\right)\Delta t,$$

and this will be zero in virtue of the last two equations.

Now the ratio between F_1 and F_2 has a simple meaning that may be derived from (29); it is related to the way in which the velocity v changes with the frequency n, that is, to the amount of the dispersion. For example, let n change by dn and let $d\left(\frac{n}{v}\right)$ be the corresponding change of $\frac{n}{v}$. Then, by (29),

$$F_1 i\, dn - F_2 i\, d\left(\frac{n}{v}\right) = 0,$$

or

$$\frac{F_1}{F_2} = \frac{d}{dn}\left(\frac{n}{v}\right), \tag{37}$$

so that (36) becomes

$$\frac{1}{w} = \frac{d}{dn}\left(\frac{n}{v}\right),$$

or

$$\frac{1}{w} = \frac{1}{v} - \frac{n}{v^2}\frac{dv}{dn}, \tag{38}$$

the formula for the group-velocity that was first given by Lord Rayleigh. If, instead of the frequency, the wave-length λ (in vacuum) is introduced,

$$\frac{1}{w}=\frac{1}{v}+\frac{\lambda}{v^2}\frac{dv}{d\lambda}. \tag{39}$$

If the last term in either of these equations is very small,

$$w=v+n\frac{dv}{dn}=v-\lambda\frac{dv}{d\lambda}. \tag{40}$$

Let μ be the index of refraction, so that

$$\mu=\frac{c}{v},\quad \frac{n}{v}=\frac{n\mu}{c};$$

then

$$\frac{1}{w}=\frac{d}{dn}\left(\frac{n\mu}{c}\right)=\frac{\mu}{c}+\frac{n}{c}\frac{d\mu}{dn}. \tag{41}$$

When $\mu=1$ we have, of course, $w=c$; and when there is no dispersion, $\frac{dv}{d\lambda}=0$ and $w=v$.

The above calculation of the group-velocity w may be considered as an application of a well-known mathematical theorem.

If x, y, z are three variables connected by one relation, we have the equation

$$\left(\frac{\partial y}{\partial z}\right)_x=-\frac{\left(\frac{\partial y}{\partial x}\right)_z}{\left(\frac{\partial z}{\partial x}\right)_y}, \tag{42}$$

which is frequently used in thermodynamics; thus when the three variables are the temperature T, the pressure p, and the specific volume v, the relation is

$$\left(\frac{\partial p}{\partial v}\right)_T=-\frac{\left(\frac{\partial p}{\partial T}\right)_v}{\left(\frac{\partial v}{\partial T}\right)_p}, \tag{43}$$

and each term has a physical meaning, the quantity on the left-hand side being the inverse of the compressibility.

Now, in order to find w we had to fix our attention on the increments that must be given to x and t in order that the amplitude a may remain constant. We may say that, by definition,

$$w = \left(\frac{\partial x}{\partial t}\right)_a.$$

By virtue of (42) we may write

$$w = -\frac{\dfrac{\partial a}{\partial t}}{\dfrac{\partial a}{\partial x}},$$

and this leads to the above result.

The formula for the group-velocity may be applied to other types of waves. We shall consider as an example the case of surface waves on deep water, a case in which there is a considerable difference between w and v.

The displacement may be expressed in the form

$$a\, e^{\frac{ny}{v}} \cos n\left(t - \frac{x}{v}\right),$$

where y is measured vertically upward and x in the direction of propagation. The velocity of propagation is given by

$$v = \frac{g}{n}.$$

Hence
$$\frac{1}{w} = \frac{d}{dn}\left(\frac{n}{v}\right) = \frac{2n}{g} = \frac{2}{v};$$

that is,
$$w = \tfrac{1}{2} v. \tag{44}$$

9. The Flow of Energy in a Beam of Light. In what precedes we have considered, in the first place, the lateral delimitation of a system of waves, and, in the second place, its delimitation by two planes which we may call the front and the rear. The two things can of course be combined.

The physical meaning both of the rays of light and of the group-velocity will perhaps be better understood if we fix our attention on the amount of electromagnetic energy and on the flow of energy that is determined by Poynting's vector. In the first place it is clear that this vector, which is perpendicular

to both **E** and **H**, being represented by $c[\mathbf{E}\cdot\mathbf{H}]$, that is, the vector product* of **E** and **H** multiplied by c, must give us the direction of the rays of light. Indeed, when we say that the vector

$$\mathbf{S}=c[\mathbf{E}\cdot\mathbf{H}] \tag{45}$$

is the flow or current of energy, we mean to assert that the amount of energy transmitted across some surface element is given (per unit of area and unit of time) by the component of **S** in the direction of the normal n to the element (Fig. 7). The flow of energy across the surface will be zero if $S_n=0$; that is, if the vector **S** has its direction in the surface. This must be true for the tubular or cylindrical surface by which a beam of light can be laterally limited, and therefore the rays of light must have the direction of Poynting's vector.

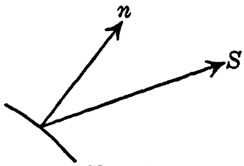

FIG. 7

Again, let us consider a beam in an isotropic medium, extending on the front side to the plane P, which is advancing with the group-velocity w. At the end of the time Δt this plane will have taken the position P', the distance between P and P' being $w\Delta t$. The energy contained in the space between P and P' is obviously equal to the energy that has flowed through P in time Δt.

Let the area of the section P of the beam be unity. As we pass along the beam the energy per unit of volume is changing continually from point to point. But if the distance between P and P' is very great in comparison with λ, we may reckon with a mean value. Let this mean value of the energy per unit of volume be U.

Similarly the flow of energy across a plane like P will not be constant; it will go on by jerks. If the time considered is very great in comparison with the period, we can reckon with a mean value, which we shall call $\bar{S}$. We now have the relation

$$\bar{S}=wU,$$

or

$$w=\frac{\bar{S}}{U}. \tag{46}$$

* This is a vector with components $E_yH_z-E_zH_y$, $E_zH_x-E_xH_z$, $E_xH_y-E_yH_x$.

10. Experimental Determination of the Velocity of Light. Which Velocity is actually Measured? We found in § 8 that in some cases (for example, in the case of surface waves on deep water) there is a great difference between the phase-velocity v and the group-velocity w. In the case of light everything depends on the amount of dispersion.

For air the difference between the two velocities is found to be not more than 1/100,000 of one of them; nevertheless it is interesting to know which of the two velocities is determined by ordinary methods.

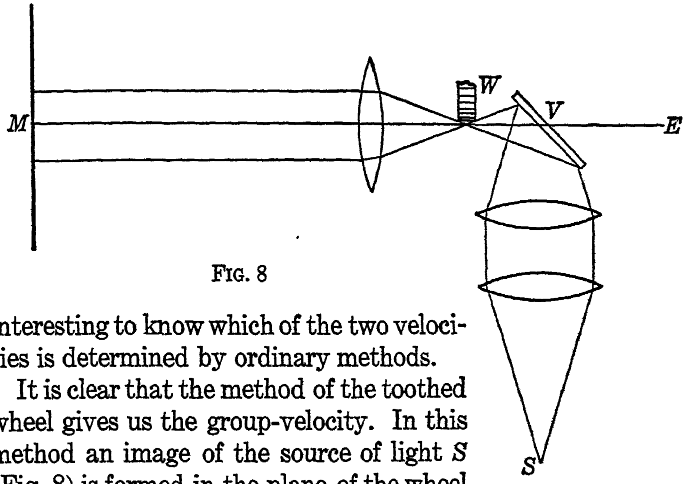

FIG. 8

It is clear that the method of the toothed wheel gives us the group-velocity. In this method an image of the source of light S (Fig. 8) is formed in the plane of the wheel W at a point on the rim, the light being focused by means of lenses and a half-silvered reflecting plate V inside a telescope. After the light has passed between the teeth of the revolving wheel it is converted into a parallel beam by the object glass of the telescope and is allowed to travel a distance of some kilometers before it is reflected back along its original path. The light is then observed through the plate V.

Whether light will be seen or not, and what will be its intensity, all depends upon the angle through which the wheel has rotated during the time required for the propagation of the light to the distant mirror and back. Let the breadth of the teeth be equal to that of the intervals between them. Under these circumstances, if, during the time in question, a tooth has

just been replaced by an interval, no light will reach the eye E. On the other hand, the intensity is a maximum when an interval has been replaced by an interval, as will be the case when the previous velocity of rotation is doubled. Now by its rotation the toothed wheel cuts off a limited portion of the incident beam, and we get a limited beam like the one considered above; it will be the propagation of these limited portions that determines the phenomena. The result of the experiment is, therefore, the group-velocity.

The theory of the other method is more difficult. This method has been improved and used with much success by

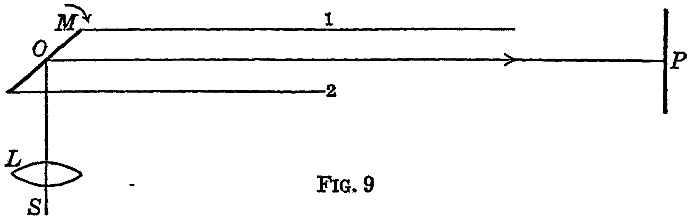

FIG. 9

Michelson, and it may be observed, in the first place, that in it also we have to do with limited trains of waves.

You know that (in one of the forms of the experiment) light coming from the collimator lens L is made to fall on the revolving mirror M. Thence it is reflected to the distant mirror P (Fig. 9). Returning along its path, it will find M in a changed position, so that it will return not to S but to a point at a short distance from S. If we rotate M slowly and receive the returning waves in the eye, we shall observe light only for short intervals of time, namely, during the intervals in which the position of the mirror M is such that the reflected rays fall on P. This remark suffices to show that we also have limited trains of waves when M is rotating at a high speed, so that things are much the same as when the toothed wheel is used.

This would lead us to expect a measure of group-velocity, but there is another circumstance to which the late Lord Rayleigh has called attention. This is the effect of Doppler's principle in the reflection of light by the moving mirror. In the case of normal reflection it is known that the frequency is changed in the

ratio of 1 to $1 \pm \frac{2v}{c}$, where v is the velocity of the mirror and c that of light. This was proved experimentally by Belopolski.

In the case of oblique incidence the term $2\,v/c$ must be multiplied by $\cos\theta$ where θ is the angle of incidence.

Now if the mirror M rotates about an axis through O perpendicular to the plane of the diagram, one half of the mirror will move toward the incident rays, and the other will recede from them. Thus in the reflected beam the frequency will be greater on the side 1 than on the side 2. If now v depends on the frequency (that is, if the beam is made to pass through a dispersive medium), the velocity of the waves will be different on the two sides. The wave-fronts will be turned, and it is easy to see that the rotations of the wave-fronts on the way from O to P and on the way back are added to each other (there is not compensation, as one might perhaps think), so that, after all, if a wave-front left the mirror M in the position α (Fig. 10), it will perhaps return to it in the position α'. Now in the reflection of light by a mirror every change in the direction of the incident waves will produce a similar change in the direction of the reflected waves, and a corresponding change in their normal, which is the direction in which the light is perceived. The observed directions would therefore be the effect of two rotations, namely, the rotation of the mirror in the period following the first reflection and the rotation of the waves themselves.

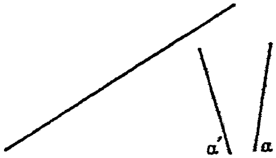

FIG. 10

This looks very satisfactory, but on going deeper into the question one encounters many difficulties. I have made some calculations, using a method similar to that which I applied to the problem of the group-velocity. In an expression of the form

$$\phi = a \cos n\left(t - \frac{\lambda x + \mu y + \nu z}{v} + p\right) \tag{47}$$

it is not only the amplitude a that can be a slowly varying function of the coördinates x, y, z and the time t, but also the

frequency n, the velocity of the waves v, and the direction cosines (λ, μ, ν) of the normal. You will see that by making suppositions of this kind we can come to exactly the same state of things that was considered by Rayleigh, and our method allows us to examine it somewhat more in detail. I shall not, however, go into these questions now.*

11. Solutions of Maxwell's Equations for the Ether. Let us now return to the ether. Maxwell's equations, which we have learned to know already, admit of many beautiful and elegant solutions. In the first place, we can eliminate from them any five of the six quantities E_x, E_y, E_z, H_x, H_y, H_z; we shall then find for the remaining one, which we shall again call ϕ, an equation of the form

$$\frac{\partial^2\phi}{\partial t^2} = c^2\Delta\phi, \tag{48}$$

where $$\Delta\phi \equiv \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}.$$

This equation reduces to our former equation when ϕ depends on one coördinate only.

Now a general solution of this equation is known. Before speaking of it we observe that if the function ϕ were to be independent of the time, we should simply have

$$\Delta\phi = 0, \tag{49}$$

Laplace's well-known equation for the potential in an electrostatic field. From it we can deduce the theorem that if around the point P as center we construct a sphere of arbitrary radius, the value of ϕ at P will be exactly the mean value of ϕ at all points of the spherical surface; that is, Σ being the surface,

$$\phi_P = \frac{1}{\Sigma}\int \phi \, d\sigma = \overline{\phi}. \tag{50}$$

From this we may deduce Earnshaw's theorem that although a small charged particle may be in equilibrium in an electrostatic field, it cannot be in stable equilibrium.

* See Note 3, Appendix.

A position of equilibrium is stable only if the potential energy is a minimum. Now in the present case the potential energy is proportional to the potential at the point P occupied by the electric charge, and so if, for a positively charged particle, the equilibrium were stable, the value of ϕ at P would be less than the value of ϕ at each point of a small sphere with P as center. This, however, is impossible, since the value of ϕ at P is the mean value of ϕ over the sphere.

As an illustration let us consider a charged particle P in equilibrium at a point midway between the centers of two equal uniformly charged spheres, both carrying a positive charge (Fig. 11).

If P carries a negative charge, the equilibrium is stable for transverse displacements but unstable for displacements along the line of centers. If, on the other hand, P carries a positive charge, the equilibrium is unstable for transverse displacements and stable for displacements along the line of centers.

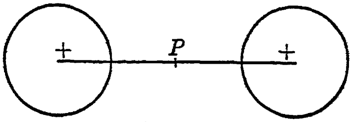

FIG. 11

This principle of Earnshaw has become of present-day interest on account of the proposed static atom and static theory of molecules and crystals. A static atom is clearly impossible if matter consists only of electric charges and these are supposed to be held together by electrostatic forces only.

If, however, the charges are in stationary motion, the principle is inapplicable and so a pure electric theory of matter might be possible. In the case of equation (48) there is a theorem similar to the one expressed by equation (50). Let us suppose that at time $t=0$ we know the values of ϕ and $\frac{\partial \phi}{\partial t}$ at each point of space, so that

$$\phi = \phi_0, \quad \frac{\partial \phi}{\partial t} = \dot{\phi}_0, \quad \text{for} \quad t = 0, \tag{51}$$

ϕ_0 and $\dot{\phi}_0$ being known functions of x, y and z. If we want to know the value of ϕ at a point P at time t, we construct around P a sphere of radius ct; then

$$\phi_P = \frac{d}{dt}(t\overline{\phi}_0) + t\overline{\dot{\phi}}_0, \tag{52}$$

where, as before, $\overline{\phi}_0$ denotes the mean value of ϕ_0 over the sphere and $\dot{\overline{\phi}}_0$ has a similar meaning. In calculating the first term on the right-hand side we must keep in mind that $\overline{\phi}_0$ is a function of t. Indeed, when t is made to increase, the radius ct of the sphere changes, so that we have to take the mean value of ϕ_0 over a new sphere.

We may deduce from this the mode of propagation of a disturbance that is at first confined to a limited space Σ (Fig. 12). The disturbance at any point P external to Σ will begin at time $t=r_1/c$, will last for a time $(r_2-r_1)/c$, and will then cease altogether at time $t=r_2/c$. The lengths r_1 and r_2 denote the radii of two spheres described with P as center, the one just excluding, the other just including, the space Σ.

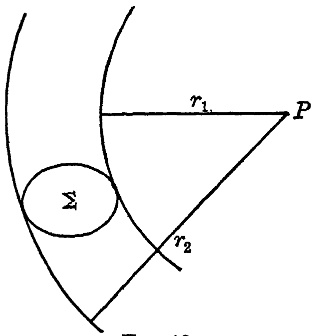

FIG. 12

This theorem, which was given by Poisson in 1807, indicates a tendency of electromagnetic disturbances to spread out in all directions. In the light of the modern theory of quanta this tendency creates difficulties, as remarked in § 6.

12. The Production of Light. As soon as it was known that light consists in a vibratory motion, it was natural to suppose that luminous bodies contain small particles vibrating about their positions of equilibrium and in this way producing waves that are propagated in the ether. Afterward, when the electromagnetic theory of light had been established, physicists were led to suppose that the vibrating particles are endowed with electric charges; indeed, this seemed to be the only way in which it could be understood that their motion gives rise to electromagnetic waves.

This has been the origin of the theory of electrons, or rather one of the lines along which we were led to it, for as a matter of fact many different lines of research have driven us to the same

conclusion. I must add that, according to the theory of electrons, not only the rays of light but all electromagnetic fields, whatever their special features, are to be considered as due to the presence and the motion of charged particles. If a conductor has a positive or a negative charge, this means that on its surface there is a distribution of positive or negative particles; each of them is surrounded by an electric field, and all these elementary fields, superposed on each other, will constitute the field around the body. If in this field we place a ponderable dielectric, there will be a change in the course of the lines of force,—a change that can be accounted for by supposing that the elements of volume are polarized, obtaining a positive charge on one side and a negative charge on the other. We have to think of this polarization as a displacement of charged particles from their original positions, the positive ones going in one direction or the negative ones in the opposite direction, or perhaps both displacements taking place at the same time.

Again, an electric current in a conducting body will be a motion of charged particles that can go on for any length of time in one and the same direction. You know that the phenomena of electrolysis led to this conclusion many years ago; the motion of the ions in a salt solution, for instance, is perfectly well known, and the effects produced by it have been examined with good success. These ions are atoms or groups of atoms. There is also good experimental evidence that metallic conductors contain freely movable charged particles, only these particles are much smaller than atoms; they are the negative electrons, whose mass is something like the 1850th part of the mass of an atom of hydrogen. Just as the electrostatic field around a charged body results from the superposition of all the elementary fields of which we spoke, so also the magnetic field around a wire that is traversed by a current will be composed of as many fields as there are movable charged particles. Indeed, such a particle, which when at rest is the center of an electric field only, will produce a magnetic field as soon as it is set in motion. That a moving charged body exerts a magnetic force and can thus give rise to a deflection of a magnetic needle

has been shown by the celebrated experiment of Rowland and by many repetitions of it, made in different forms.

Now, if our equations are to express not only the modes of distribution and propagation of electromagnetic fields but also the way in which they are produced, we must include in them the electric charges. This is done by a slight extension, to which we shall now proceed.

13. The Field Equations of the Electron Theory. We shall suppose the electric charges to be distributed over space with a certain density ρ, so that ρ means the electric charge per unit of volume. We shall further assume that ρ is a continuous function of the coördinates x, y, z, so that there is no sharp boundary; if a charge is limited to a certain finite region, there will be all around that region a transition layer, of a certain thickness, in which ρ changes gradually from the value it has on the inside boundary to zero. This assumption, however, is introduced solely for the sake of mathematical convenience. We can choose the thickness of the layer of transition as small as we like, and so, by passing to the limit, deduce from our results those that will hold in the case of a sharp boundary. By a similar process we can even come to the case of a charge distributed not over a space but over a surface; we first imagine a distribution in a layer between two surfaces σ_1 and σ_2, and then make these approach one another indefinitely, the volume-density increasing in such a manner that the amount of charge per unit area of σ remains finite.

The first change is to be made in the equation

$$\operatorname{div} \mathbf{E} = 0,$$

which is now to be replaced by

$$\operatorname{div} \mathbf{E} = \rho. \tag{53}$$

In order to understand this you must remember that the vector $\mathbf{E}$ represents not only the electric force in the ether but also the dielectric displacement. If for a moment we think of electricity as an incompressible fluid, it is clear that whenever a body P in some way or other gets a greater amount of the

fluid than is its due while in the natural state, an equal amount of the fluid will be displaced outward through any closed surface surrounding the body. If e is the charge of the body, this is expressed by the formula

$$\int E_n \, d\sigma = e, \tag{54}$$

where E_n is the component of $\mathbf{E}$ along the normal $\mathbf{n}$ to the surface, $\mathbf{n}$ being directed toward the outside of the surface. The formula is also true if e is negative.

Let us apply the last equation to the infinitely small parallelepiped of volume $dx\,dy\,dz$ (Fig. 13); then $e = \rho\,dx\,dy\,dz$. Since this is infinitely small of the third order, we must calculate the surface-integral in (54) up to the same order, and we must therefore take into account the changes of the components of $\mathbf{E}$ from one point of the surface of the parallelepiped to the other, these changes being of the first order. Now the change of E_x when we pass from a point A in the plane of area $dy\,dz$ on the left-hand side to the corresponding point in the opposite face is

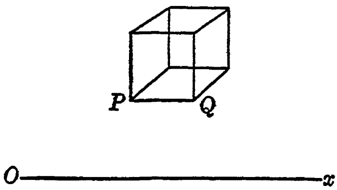

FIG. 13

$$\frac{\partial E_x}{\partial x}\,dx,$$

where for the differential coefficient we may take its value at P. From this it is easily seen that the two planes $dy\,dz$ contribute to the integral the quantity

$$\frac{\partial E_x}{\partial x}\,dx\,dy\,dz,$$

and similarly for the pairs of planes perpendicular to Oy and Oz respectively. The final result is

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho,$$

which is our equation (53). This equation may be regarded as the definition of ρ, and this way of defining the density of

electric charge is very natural. We infer the existence of a charge on a body by observing the electric forces to which it gives rise in the surrounding space, and so from the electric forces existing at the surface of a small parallelepiped we infer that there is a charge inside the surface.

It should be noticed that the vector **E** is no longer solenoidally distributed when there are electric charges.

The expression div **E** was termed by Maxwell the divergence of the vector **E** because when it is positive at a point P the vector in the immediate neighborhood of P is directed somewhat more away from P than toward it. The equation

$$\text{curl}\,\mathbf{H} = \frac{1}{c}\,\dot{\mathbf{E}} \tag{55}$$

must likewise be modified. The general form is

$$\text{curl}\,\mathbf{H} = \frac{1}{c}\,\mathbf{C}, \tag{56}$$

where **C** is the electric current. Now

$$\mathbf{C} = \dot{\mathbf{E}} + \rho\mathbf{v}, \tag{57}$$

where **v** is the velocity of the charge. The current **C** thus consists of two parts, a displacement current $\dot{\mathbf{E}}$ and a convection current $\rho\mathbf{v}$. The last-named current satisfies the equation

$$\text{div}\,(\rho\mathbf{v}) + \frac{\partial\rho}{\partial t} = 0, \tag{58}$$

which can easily be proved (it is the exact analogue of the equation of continuity in hydrodynamics), and by means of which, in connection with (53), it can be shown that the current **C** given by (57) is solenoidally distributed, — in other words, that div $\mathbf{C} = 0$.

It may be well to insert a few words here on the meaning of the curl of a vector, or, as it is also called, the rotation of a vector.

Let **A** be the vector and consider its line-integral along a closed curve C,

$$L = \int_C A_s\,ds,$$

A_s denoting the component of **A** in the direction in which the arc s increases. If **A** is a force, the line-integral represents the

work done by the force in a circuital displacement, and this is not generally zero. We can, in fact, prove the following theorems:

1. If C is an infinitely small closed plane curve near the point P, the value of L is proportional to the inclosed area, depending only on the position of P and the direction of the plane, not on the form of the closed curve.

2. If C is in a plane parallel to YOZ, s increasing in a direction corresponding to a rotation through 90° from OY to OZ, the value of the integral is

$$\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) d\sigma,$$

where $d\sigma$ is the inclosed area. This may be proved by considering a small rectangle of sides dy, dz and calculating the line-integral up to quantities of the second order. The change in the value of A_z when we pass from a point on the left-hand side dz in the figure to the corresponding point on the opposite side is

$$\frac{\partial A_z}{\partial y} dy,$$

and therefore the two sides dz contribute together to the integral, paying due regard to sign, the quantity

$$\frac{\partial A_z}{\partial y} dy\, dz.$$

The contribution of the two sides dy is found in a similar way, and so we find for the line-integral the value stated, since $d\sigma = dy\, dz$.

In planes perpendicular to OY and OZ, $d\sigma$ must be multiplied by

$$\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \quad \text{and} \quad \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}$$

respectively. Let **B** be the vector with the components

$$\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \quad \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}, \quad \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y};$$

then in each of these three cases $d\sigma$ must be multiplied by a component of **B**.

3. If once we know the values of the line-integral L for infinitely small closed circuits in planes at right angles to the axes of coördinates, we can deduce from them the value of L for a closed curve C lying in any plane.

Let PQ, PR, PS (Fig. 14) have the directions of the axes of coördinates, QRS being a plane of any orientation. The line-integral along the path QRS is equal to the sum of the line-integrals along PQR, PRS, PSQ. The result is that the line-integral along QRS is $B_n\, d\sigma$, where $d\sigma$ is the area QRS and n the normal to the plane QRS in a direction corresponding to

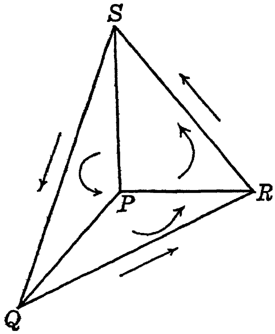

FIG. 14

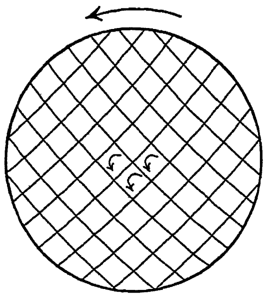

FIG. 15

the direction chosen as positive along the boundary (in this case QRS). This result may now be extended to any small area $d\sigma$ and finally to any superficial area which is divided up into small circuits which are approximately plane (Fig. 15).

Thus the vector $\mathbf{B}$ determines the value of the line-integral of $\mathbf{A}$ for any closed curve. It is called the rotation, or curl, of the vector $\mathbf{A}$, because when $\mathbf{B}$ is different from zero there is something like a turning or whirling in a definite direction in the vector field of $\mathbf{A}$.

If $\mathbf{B}$ is everywhere zero, the vector $\mathbf{A}$ is said to be irrotationally distributed. This occurs when the vector $\mathbf{A}$ can be derived from a potential ϕ; that is, when

$$A_x = -\frac{\partial \phi}{\partial x}, \text{ etc.,}$$

or, in vector language, $\mathbf{A} = -\operatorname{grad} \phi$. (59)

It is well known that in a field of force with a one-valued potential (that is, a conservative field of force) the work in a closed circuit is zero. That in this case curl $\mathbf{A} = 0$ appears also from the formula. For, by (59),

$$\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0, \text{ etc.}$$

To the equations

$$\operatorname{div} \mathbf{E} = \rho, \quad \operatorname{curl} \mathbf{H} = \frac{1}{c}(\dot{\mathbf{E}} + \rho \mathbf{v}) \tag{60}$$

we must next add

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{curl} \mathbf{E} = -\frac{1}{c}\dot{\mathbf{H}} \tag{61}$$

without any alteration.

A solution of these equations that is often very useful is obtained by means of two potentials, a scalar potential ϕ and a vector potential $\mathbf{A}$. By way of introduction we recall one or two known theorems.

1. In a stationary field, such as an electrostatic field,

$$\dot{\mathbf{H}} = 0 \quad \text{and} \quad \operatorname{curl} \mathbf{E} = 0; \tag{62}$$

hence the electric force can be derived from a potential ϕ as follows:

$$\mathbf{E} = -\operatorname{grad} \phi, \tag{63}$$

ϕ being the electrostatic potential.

2. If

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \text{ etc.,}$$

we have div $\mathbf{B} = 0$. The curl of a vector is thus solenoidally distributed, and the converse proposition is equally true. A vector solenoidally distributed can always be considered as the curl of another vector. We have seen already that in all cases div $\mathbf{B} = 0$, $\mathbf{B}$ being the magnetic induction. Therefore we can always introduce a vector $\mathbf{A}$ from which the magnetic induction can be derived by means of the equation

$$\mathbf{B} = \operatorname{curl} \mathbf{A}. \tag{64}$$

$\mathbf{A}$ is called the vector potential. In our case $\mathbf{B} = \mathbf{H}$, so that the magnetic force can be derived from a vector potential.

In the general case there is a scalar potential ϕ and a vector potential $\mathbf{A}$.

From these the electric and magnetic forces are derived as follows:

$$\mathbf{E} = -\operatorname{grad}\phi - \frac{1}{c}\dot{\mathbf{A}}, \quad \mathbf{H} = \operatorname{curl}\mathbf{A}. \tag{65}$$

The two potentials are connected by the relation

$$\operatorname{div}\mathbf{A} = -\frac{1}{c}\dot{\phi} \tag{66}$$

and satisfy the two partial differential equations

$$\Box\phi = -\rho, \quad \Box\mathbf{A} = -\frac{1}{c}\rho\mathbf{v}, \tag{67}$$

where

$$\Box\phi \equiv \Delta\phi - \frac{1}{c^2}\ddot{\phi} \tag{68}$$

and $\Box\mathbf{A}$ denotes the vector with components of type $\Box A_x$.

If the state were stationary, so that $\ddot{\phi} = 0$, $\ddot{\mathbf{A}} = 0$, these equations would have the form of Poisson's equation $\Delta\phi = -\rho$. If, on the other hand, the right-hand sides were zero, we should have the equations of propagation similar to those considered previously. The influence of the charges makes itself felt in the terms on the right-hand sides of (67); it is these terms that will lead to a production of vibrations if, for instance, the charges have a periodic motion so that at a definite point of space the values of ρ, ρv_x, ρv_y, ρv_z change from one moment to another.

A solution of these equations may be obtained by a generalization of a theorem given by Kirchhoff; it takes the form

$$\left.\begin{aligned} \phi &= \frac{1}{4\pi}\int\frac{1}{r}[\rho]\,dV \\ \mathbf{A} &= \frac{1}{4\pi c}\int\frac{1}{r}[\rho\mathbf{v}]\,dV \end{aligned}\right\}, \tag{69}$$

where dV is an element of volume. The square brackets indicate that the quantity inclosed is to be taken for the time $t - r/c$.

By these equations, combined with (65), our problem is solved. They show that in order to calculate the field we have to proceed as follows:

Let P be the point at which we wish to determine the potentials at time t. We must divide the whole surrounding space into elements of volume, any one of which is called dV. Let it be situated at Q and let $QP = r$. In this element of space there may or may not be a part of an electron at a certain time. We are concerned only with the question whether it contains a charge at time $t - r/c$. Indeed, the brackets serve to remind us that we are to understand by ρ the density existing in dV at time $t - r/c$ and by $\rho\mathbf{v}$ the product of this density and the velocity of the charge within dV at that same instant. These values $[\rho]$ and $[\rho\mathbf{v}]$ must be multiplied by dV and divided by r. Finally, we have to do for all the elements what we have done for the one dV and to add the results.

For a now obvious reason $\mathbf{A}$ and ϕ are called retarded potentials. We have now found the means of determining the field (always according to classical theory) that is due to a given distribution and motion of electric charges.

14. Applications of the Foregoing Formulæ. Consider an electron, that is, a charged particle with a certain distribution of charge of volume-density ρ, to have an oscillatory motion in the direction of the axis of x, the displacement from the position of equilibrium being $f(t)$ at time t.

The quantity f is to be considered as very small, so that squares and higher powers of f may be neglected in the expressions for the scalar and the vector potential, and also in the expressions for the electric and the magnetic force.

We want to calculate ϕ and $\mathbf{A}$ for a point P at time t. In doing so we must keep in mind one thing that could easily be overlooked at first sight; namely, we have to integrate over all the places in space at which at the time $t - r/c$ there is an electric charge, and this is not the same as an integration over the space occupied by the electron at some definite instant.

In order to make this clear we shall fix our attention on some point Q of the electron. Among the different positions that it takes in the course of time there is just one, Q_e (Fig. 16), satisfying the condition (and determined by it) that if it were the center of some disturbance traveling onward with the speed c, this disturbance would reach the point P exactly at the time t. We shall call Q_e the effective position of the point Q under consideration. Now consider a number of points at infinitely small distances around Q; let R_e, S_e, $\cdots$ be their effective positions, whereas the positions Q, R, S, etc. are simultaneous. Then the element of space which contains these latter positions will be an element of volume of the charged particle, and Q_e, R_e, S_e, $\cdots$ will determine an "effective" element of volume, such as we have to take for dV in our formula. This element will have to be multiplied by the density ρ existing at the point Q_e.

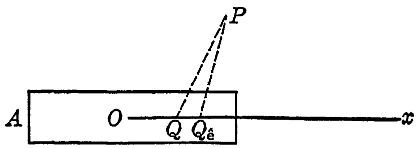

FIG. 16

Let AB be an infinitely thin cylinder in the charged particle of normal section ω and having its length in the direction of Ox. I mean to say that AB represents this cylinder in its position of equilibrium. We shall choose as origin of coördinates a definite point O of the cylinder, likewise in its position of equilibrium, and we shall denote by Q the position of equilibrium of some other point of the cylinder. We shall write ξ for the relative coördinate OQ, so that the coördinates of Q are

$$(\xi, 0, 0).$$

Now we must fix our attention on the effective position Q_e of Q. This will be the position of the point at the time $t - \frac{1}{c}(PQ_e)$; but since the amplitude and velocity of the vibration are infinitely small, and as we shall neglect in the potentials quantities of the second order, we may just as well say that Q_e is the position of Q at the time $t - \frac{1}{c}(PQ)$. Hence, if we write r

for the distance of P from Q, that is, from the position of equilibrium of the point considered, the coördinates of Q_e will be

$$\xi + f\left(t - \frac{r}{c}\right),\ 0,\ 0.$$

Here r is a function of ξ. Therefore, if, for a definite value of t, we pass from Q to a neighboring point R, the distance QR being $d\xi$, the distance between the effective positions Q_e and R_e will be

$$\left[1 + \frac{\partial}{\partial \xi} f\left(t - \frac{r}{c}\right)\right] d\xi.$$

We can now take, for the effective element of volume,

$$dV = \left[1 + \frac{\partial}{\partial \xi} f\left(t - \frac{r}{c}\right)\right] \omega\, d\xi.$$

In order to calculate $\phi = \frac{1}{4\pi}\int \frac{[\rho]\, dV}{r}$

we must multiply dV by the density ρ existing at the point Q_e and by $1/r_e$; we must also take into account the difference between $1/r_e$ and $1/r$. This difference is given by

$$\frac{1}{r_e} = \frac{1}{r} + f\left(t - \frac{r}{c}\right)\frac{\partial}{\partial \xi}\left(\frac{1}{r}\right),$$

so that we find

$$\frac{dV}{r_e} = \left\{\frac{1}{r} + \frac{\partial}{\partial \xi}\left[\frac{1}{r} f\left(t - \frac{r}{c}\right)\right]\right\} \omega\, d\xi.$$

As the density at the point considered is the same, whether its position be Q or Q_e, we have

$$\phi = \frac{1}{4\pi}\int \rho \left\{\frac{1}{r} + \frac{\partial}{\partial \xi}\left[\frac{1}{r} f\left(t - \frac{r}{c}\right)\right]\right\} \omega\, d\xi.$$

By this we see that the scalar potential, which is due to our infinitely thin cylinder, consists of two parts. We shall find corresponding parts if, by an integration over all the cylinders into which the charged particle can be divided, we pass on to the value of ϕ due to the whole particle.

The first part of ϕ will be

$$\frac{1}{4\pi}\int \frac{\rho}{r}\, \omega\, d\xi = \frac{1}{4\pi}\int \frac{\rho}{r}\, dV.$$

This is simply the electrostatic potential that would exist if the particle were constantly at rest.

Remembering that, when we operate on a function of r, if x, y, z are the coördinates of the point P for which we want to calculate the potentials,

$$\frac{\partial}{\partial \xi} = -\frac{\partial}{\partial x},$$

we can write for the second part

$$-\frac{1}{4\pi}\frac{\partial}{\partial x}\int \frac{\rho}{r} f\left(t-\frac{r}{c}\right)\omega d\xi.$$

We can simplify this result by supposing that the dimensions of the charged particle are extremely small in comparison with the distance OP and also with respect to the distance over which a disturbance traveling with the velocity c would be propagated in an interval of time during which $f(t)$ changes appreciably. Then, in all elements of our integral, we can understand by r the distance of P from the center of the particle in its position of equilibrium, so that we have simply, for the second part of the scalar potential,

$$-\frac{1}{4\pi}\frac{\partial}{\partial x}\left[\frac{e}{r} f\left(t-\frac{r}{c}\right)\right].$$

In calculating the vector potential we notice that of the components of $\mathbf{v}$ only the first, namely, v_x, is different from zero; therefore $A_y = 0$, $A_z = 0$. To find the value of A_x we remark that we have to introduce the factor

$$v_x = f'\left(t-\frac{r}{c}\right)$$

into the preceding formula for ϕ; but in making the change we can confine ourselves to the first part of ϕ, because both f and f' are infinitely small, so that the product can be neglected. We finally obtain

$$A_x = \frac{e}{4\pi c r} f'\left(t-\frac{r}{c}\right). \tag{70}$$

In our determination of the field we shall omit the electrostatic part; namely, the part that depends on the first part of ϕ. Indeed, this electrostatic field may be made to disappear by the supposition that in the immediate neighborhood of our moving charged particle there are one or more others that are at rest and produce an electrostatic field equal and opposite to the one of which we have just spoken. However this may be, if we want only to know the vibrations that are produced, and not the electrostatic field on which these vibrations are eventually superposed, we can put

$$\phi = -\frac{e}{4\pi}\frac{\partial}{\partial x}\left[\frac{1}{r}f\left(t-\frac{r}{c}\right)\right], \tag{71}$$

taken together with the above value of A_x.

In the case of simple harmonic vibrations we may write

$$f(t) = a\cos n(t+p),$$

where n determines the frequency and p the phase. The expressions for the potentials are now

$$\phi = -\frac{\partial G}{\partial x}, \quad A_x = \frac{1}{c}\frac{\partial G}{\partial t}, \quad A_y = 0, \quad A_z = 0, \tag{72}$$

where
$$G = \frac{ae}{4\pi r}\cos n\left(t-\frac{r}{c}+p\right). \tag{73}$$

The components of the electric and magnetic forces are

$$\left.\begin{aligned} E_x &= -\frac{\partial\phi}{\partial x} - \frac{1}{c}\frac{\partial A_x}{\partial t} = \frac{\partial^2 G}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 G}{\partial t^2}, \quad H_x = 0 \\ E_y &= -\frac{\partial\phi}{\partial y} = \frac{\partial^2 G}{\partial x\,\partial y}, \quad H_y = \frac{\partial A_x}{\partial z} = \frac{1}{c}\frac{\partial^2 G}{\partial z\,\partial t} \\ E_z &= -\frac{\partial\phi}{\partial z} = \frac{\partial^2 G}{\partial x\,\partial z}, \quad H_z = -\frac{\partial A_x}{\partial y} = -\frac{1}{c}\frac{\partial^2 G}{\partial y\,\partial t} \end{aligned}\right\}. \tag{74}$$

It is easily verified that Maxwell's equations (for the ether) are satisfied by these expressions, because the function G, and more generally the function

$$\psi = \frac{1}{r}f\left(t-\frac{r}{c}\right),$$

satisfies $\Box\psi = 0$.

At a great distance from the vibrating particle the most important terms are those of order $1/r$; these are

$$\left.\begin{aligned} E_x^0 &= \frac{n^2}{c^2}\left(1-\frac{x^2}{r^2}\right)G, & H_x^0 &= 0 \\ E_y^0 &= -\frac{n^2}{c^2}\frac{xy}{r^2}G, & H_y^0 &= \frac{n^2}{c^2}\frac{z}{r}G \\ E_z^0 &= -\frac{n^2}{c^2}\frac{xz}{r^2}G, & H_z^0 &= -\frac{n^2}{c^2}\frac{y}{r}G \end{aligned}\right\}. \tag{75}$$

15. Interpretation of These Equations. As was to be expected, the field is symmetrical around the axis Ox, along which the electron is vibrating, the lines of electric force lying in planes passing through Ox and the lines of magnetic force being circles around this axis.

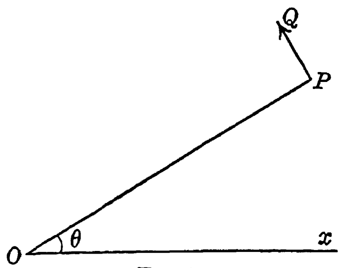

FIG. 17

At a point P both $\mathbf{E}^0$ and $\mathbf{H}^0$ are perpendicular to OP. If for the positive directions we take PQ (Fig. 17) (corresponding to an increase of θ) and a line perpendicular to POx (corresponding to a rotation from Ox to OP), we have

$$|\mathbf{E}^0| = |\mathbf{H}^0| = -\frac{aen^2}{4\pi c^2 r}\sin\theta\cos n\left(t-\frac{r}{c}+p\right),$$

where $\mathbf{E}^0$, when this expression is positive, is in the direction of the first line just mentioned, and $\mathbf{H}^0$ is in the direction of the second.

The result is similar to the one found for the case of a plane polarized beam of parallel rays of light.

The flow of energy, which is determined by Poynting's vector $\mathbf{S}^0 = c\,[\mathbf{E}^0 \cdot \mathbf{H}^0]$, is directed along OP prolonged and has the intensity

$$|\mathbf{S}^0| = c\,|\mathbf{E}^0|\,|\mathbf{H}^0| = \frac{a^2e^2n^4}{16\pi^2c^3r^2}\sin^2\theta\cos^2 n\left(t-\frac{r}{c}+p\right). \tag{76}$$

We shall consider its mean value over a full period (or over a long lapse of time). The mean value of $\cos^2 n\left(t - \frac{r}{c} + p\right)$ is $\frac{1}{2}$, so that we get

$$|\overline{\mathrm{S}^0}| = \frac{a^2e^2n^4}{32\,\pi^2c^3r^2}\sin^2\theta. \tag{77}$$

We may deduce from this expression the amount of energy radiated by a vibrating particle per unit of time. Let us draw a sphere of radius r around the point O and integrate $|\overline{\mathrm{S}^0}|$ over its surface. If $d\sigma$ is an element of surface, this amount of energy is given by

$$\frac{a^2e^2n^4}{32\,\pi^2c^3r^2}\int \sin^2\theta\,d\sigma.$$

We have here to replace $\sin^2\theta$ by its mean value over the sphere. Now the mean value of $\cos^2\theta$ is $\frac{1}{3}$; indeed, the mean values of the squares of the three direction constants of the radius are equal by symmetry, and the mean value of their sum is unity. The mean value of $\sin^2\theta = 1 - \cos^2\theta$ is consequently $\frac{2}{3}$. Hence, since

$$\int d\sigma = 4\,\pi r^2,$$

our last expression becomes $\dfrac{a^2e^2n^4}{12\,\pi c^3}$. (78)

16. Nature of the Field at a Small Distance from the Vibrating Electron. Let us now suppose that the distance r is small in comparison with the wave-length but still large in comparison with the dimensions of the electron.

In the former case we could confine the differentiations with respect to the coördinates to the argument of the cosine and could consider as constant the amplitudes by which

$$\cos n(t - \frac{r}{c} + p)$$

is multiplied. It is easily seen that this is permissible so long as $r >> \lambda$.

When, on the other hand, $r << \lambda$, we have only to differentiate amplitudes and can omit the differentiation with respect

to t, at the same time replacing $\cos n(t - \frac{r}{c} + p)$ by $\cos n(t + p)$. We then get

$$\mathbf{E} = \frac{ae}{4\pi} \cos n(t + p) \text{ grad } \left[\frac{\partial}{\partial x}\left(\frac{1}{r}\right)\right], \qquad \mathbf{H} = 0. \tag{79}$$

Adding these expressions to those deduced from the previously omitted first part of ϕ, we infer that at short distances electrostatic forces predominate, these being calculated as if the particle were at rest in its instantaneous position. The terms which we first considered, and which prevail at great distances, can be called radiation terms. The amplitude for these terms is inversely proportional to the first power of the distance, and it is precisely on this account that Poynting's vector, that is, the flow of energy through unit area, is inversely proportional to r^2; this is the well-known elementary law.

It must of course be true that during a full period as much energy passes through one sphere around the luminous point as through another, because the state of things in the space between them is the same at the beginning as at the end of the period.

If we construct around the luminous point as center a sphere with radius r, neither very great nor very small in comparison with λ, so that no terms in our equations can be neglected, and if then we calculate the flow of energy through the sphere, we shall again be led to the result (78).

17. The Field of an Electron which is moving in an Arbitrary Manner with Velocity less than c. The method that served us for the vibrating particle can be used also for a particle having any given motion of translation, which will generally be with varying velocity along a curved line.

At time t let us choose a point P whose distance from the electron is large in comparison with the electron's linear dimensions. Let r be the distance from the electron to P, $\mathbf{v}$ the velocity, and v_r its component in the direction of r. Then we have

$$\phi = \frac{e}{4\pi\left[r\left(1 - \frac{v_r}{c}\right)\right]}, \qquad \mathbf{A} = \frac{e[\mathbf{v}]}{4\pi c\left[r\left(1 - \frac{v_r}{c}\right)\right]}, \tag{80}$$

where the square brackets indicate that we must fix our attention on the effective position M; by r we have to understand the distance MP, by $\mathbf{v}$ the velocity in the effective position, and by v_r its component along MP. From these formulæ the field can again be derived. If we consider only velocities and accelerations that are not too great, so that products and squares of these quantities can be neglected, it is found that at a great distance from the electron the dominant terms in the expressions for the electric and magnetic forces in the field are determined by the acceleration of the electron. The result is as follows:

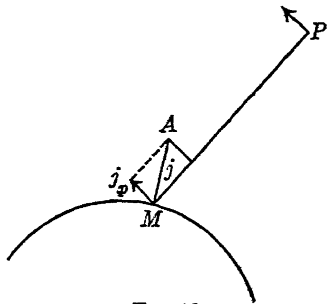

FIG. 18

Let $\dot{\mathbf{j}} = MA$ (Fig. 18) be the acceleration of the electron in its effective position, and let us resolve it into a component along MP and a component $\dot{j}_p$, at right angles to that line. Then the electric force at P will have the direction of $\dot{j}_p$ and the magnitude

$$-\frac{e}{4\pi c^2 r}\dot{j}_p.$$

The magnetic force at P is found to be of the same magnitude and to be at right angles to the plane PMA, its direction being such that the flow of energy is in the direction of MP prolonged. The intensity of this flow per unit area of a surface perpendicular to MP is given by

$$\frac{e^2}{16\pi^2 c^3 r^2}\dot{j}_p^2 = \frac{e^2}{16\pi^2 c^3 r^2}\sin^2\theta \cdot \dot{j}^2,$$

where θ is the angle between MP and the direction MA of the acceleration. Now, if we consider a sphere around M with radius MP, M is the effective position with respect to all points of the surface, so that our last expression, with the same $\dot{j}$, may be applied all over the surface. As the mean value of $\sin^2\theta$ is $\frac{2}{3}$, the flow of energy through the sphere is found to be

$$\frac{e^2}{6\pi c^3}\dot{j}^2. \qquad (81)$$

It is important to notice that so long as the electron has a velocity which is constant in magnitude and direction, there is no radiation. But there is a radiation of energy in all cases in which there is an acceleration; that is, in all cases in which the velocity changes either in magnitude or in direction.

Formula (78) is a particular case of (81). Indeed, if the electron is vibrating in such a manner that the displacement from the position of equilibrium at time t is

$$a \cos n(t + p),$$

we have

$$j = - an^2 \cos n\left(t - \frac{r}{c} + p\right),$$

so that (81) becomes

$$\frac{a^2e^2n^4}{6\,\pi c^3} \cos^2 n\left(t - \frac{r}{c} + p\right),$$

the mean value of which over a complete period is *

$$\frac{a^2e^2n^4}{12\,\pi c^3}.$$

18. The Reaction of Radiation on a Moving Electron. Closely connected with the radiation of energy by a moving electron is a certain force that is exerted on it by the electromagnetic field and that may be called a resistance. The case is exactly similar to that of a body vibrating in the air (for instance, a tuning fork) and producing sound waves. In these waves, just as in our electromagnetic waves, a certain amount of energy is propagated outward. Hence, unless the vibrations of the body are maintained by a properly regulated external force, the amplitude of the vibrating body must gradually diminish, the body losing the energy that is radiated. We can calculate this resistance by means of the differential equations that determine the motion of the air, and in exactly the same way we can calculate the radiation resistance in the case of our electron by means of the equations of the electromagnetic field.

* For some further remarks see Note 4, Appendix.

Let me recall to you that the force with which the field acts on an electric charge is composed of two parts. Per unit of charge, the first part is given by **E** and the second by $\frac{1}{c}[\mathbf{v}\cdot\mathbf{H}]$, where the last factor denotes the vector product of the velocity **v** of the charge and the magnetic field. The total force on a charged particle can be found by multiplying the above quantities by $\rho\, dV$ (the charge on an element of volume) and by integrating the result over the space occupied by the particle.

The force $\frac{1}{c}[\mathbf{v}\cdot\mathbf{H}]$ is the one that produces the well-known deflection of cathode rays, or of β rays in a magnetic field. To this force we can also ascribe the action of a magnetic field on a metallic wire through which a current is made to pass.

Now it is particularly interesting that an electron experiences a force that is due to its own field. The calculation of the field in the interior of the electron, which, like the outside field, depends on our two potentials, shows that there are two forces, one that is proportional to the acceleration of the particle and directed oppositely to it, and another that is proportional to the differential coefficient of the acceleration with respect to the time.

The first force can be represented by $-mj$, where j is the acceleration and m a constant that has the value

$$m = \frac{e^2}{6\,\pi c^2 R} \tag{82}$$

the surrounding fluid will not be constant all over its surface, and this will lead to a resultant force oppositely directed to the acceleration j and proportional to it. If this force is $-mj$, the factor m is what we might call the "hydrodynamic" mass. We can also say that if the sphere is set in motion, a certain momentum must be given not only to the body itself but also to the surrounding water, so that, for the same acceleration, a greater force will be required than if there were no water.

We can express the same thing by saying that the presence of the water has an effect similar to that of an increase of mass.

But I shall not now discuss at any length this question of electromagnetic mass. It may suffice to remark that what we have said is true only so long as the velocity of the particle is small compared with the speed of light.

The second force is also proportional to the square of the charge, but it is independent of the shape and size of the electron; the second force is always represented by *

$$\frac{e^2}{6\,\pi c^3}\frac{d}{dt}(\mathrm{j}) = \frac{e^2}{6\,\pi c^3}\,\ddot{\mathrm{v}}. \tag{83}$$

In the case of a particle performing a simple harmonic vibration of frequency n we have $\ddot{\mathrm{v}} = -n^2\mathrm{v}$, so that the expression for the force becomes

$$-\frac{n^2e^2}{6\,\pi c^3}\,\mathrm{v},$$

showing that it is continually opposite to the velocity; the force can then truly be called a resistance.

be equal to the energy radiated, calculated in the same way. The work done by the force per unit time is, in fact,

$$v \cdot \frac{n^2 e^2}{6 \pi c^3} v = \frac{n^2 e^2 v^2}{6 \pi c^3} = \frac{n^4 e^2 a^2}{6 \pi c^3} \sin^2 (nt + p),$$

the mean value of which over a period is

$$\frac{n^4 e^2 a^2}{12 \pi c^3},$$

which agrees with (78).

There is a close analogy, which I need scarcely point out, between the vibrations excited by the moving electron and the waves emitted by an antenna in wireless telegraphy. I can now add that in the case of the antenna also we can speak of a radiation resistance that must be taken into account when we have to design a wireless installation. The expression for this resistance may be found in a similar way.

19. Difficulties in the Theory of Radiation. The classical theory of electrons has led us to the important conclusion that an electron can never have its velocity changed, either in direction or in magnitude, without becoming a center of radiation. By this the first great discrepancy between the old and the new theories shows itself. Indeed, as you know, the conception of Rutherford and Bohr as to the structure of atoms, namely, the hypothesis that an atom consists of a positive nucleus with electrons revolving around it, has proved most fertile for the

What we want to know is this: Are we wholly on the wrong track when we consider changes in velocity as accompanied by a radiation, or is there at least some truth in the idea? That electrons radiate when they have any other than a uniform rectilinear motion has never been proved experimentally, and, according to the theoretical formula, the effects which we could expect are so small that it will be scarcely possible to observe them; but there are two groups of phenomena which seem to show that a radiation by moving electrons really exists.

In the first place, we can certainly say that the waves with which we operate in wireless telegraphy are produced by rapidly alternating currents in conductors or systems of conductors, — for instance, in an antenna. Therefore, if it is true that a current in a metal consists in a motion of negative electrons (and we can infer this from the experiment of Tolman and Stewart), here is a case of radiation by moving electrons.

20. The Experiments of Tolman and Stewart. We shall now make a short digression in order to discuss these experiments, which are really very interesting. They have been published in the *Physical Review*, 2d series, Vol. 8 (1916), p. 97.

By way of introduction we shall consider a closed annular tube of glass having the form of a circle and free to rotate about its geometrical axis. Let this tube be filled with water and let it be set in motion; after a time it will have acquired a certain velocity, which is then kept constant. On account of the friction the water will in time move with this same velocity, so that eventually there will be no relative motion of the water with respect to the tube; but in the first few moments after

Now we can get an idea of the experiment made by Tolman and Stewart if we substitute for the tube a circular metallic wire and for the water the free charged particles which are contained in it and whose motion relative to the wire constitutes the electric current.

When the wire is set in motion (I mean, when a rotation about its geometric axis is imparted to it), a transient current will be generated, — transient because the forces between the metal and the free particles, which can be compared to the friction in our former case, and to which the resistance of the wire is due, will finally make the free particles move with the velocity of the wire. There is then no longer any current, because the positively and negatively electrified particles will be moving with the same velocity. When the wire is brought to rest there will be a second transient current in a direction opposite to the first.

The phenomena in question can be very easily observed with water in a tube (we may, in fact, substitute a tumbler for the tube), but the electrical experiment is a delicate one, and its successful performance is to be considered as a great achievement. But let me first give you the theory.

1. Consider a circular tube whose section is very small compared to the inclosed area, so that all points of the tube may be said to move with the same velocity w, the motion being a rotation round the geometrical axis of the tube. Let v be the velocity of the water relative to the tube, m the mass of the water per unit length of the tube, and α a constant determining the resistance. The equation of motion of the water is then

$$m \frac{d}{dt}(v + w) = -\alpha v.$$

If w is a given function of the time t, this equation can be integrated with the aid of the integrating factor $e^{\frac{\alpha t}{m}}$, but such an integration is not necessary for our purpose. Writing the equation in the form

$$v = -\frac{m}{\alpha}\frac{d}{dt}(v + w)$$

and integrating between t_1 and t_2, we conclude that the length of the column of water which flows across a section of the tube in the interval between t_1 and t_2 is

$$\int_{t_1}^{t_2} v\,dt = \frac{m}{\alpha}[(v_1 + w_1) - (v_2 + w_2)].$$

In the special case in which we have initially $w_1 = 0$, $v_1 = 0$ and finally $w_2 = w$, $v_2 = 0$, the right-hand side becomes

$$-\frac{m}{\alpha}w$$

and gives the amount of the lag of which we spoke.

We get the opposite phenomenon when we have initially $w_1 = w$, $v_1 = 0$ and finally $w_2 = 0$, $v_2 = 0$.

2. In the case of the metallic wire the theory is somewhat more complicated on account of self-induction.

Let the velocity of the wire be w and the relative velocity of the electrons v. The electric current is then

$$i = Nev,$$

where e is the charge on our free particle or electron and N is the number of such particles per unit length of the wire.

Let L be the self-induction and r the resistance of the wire in the ordinary sense. This means that the line-integral of electrical force along the wire has the value

$$-L\frac{di}{dt}$$

and that there is a force, similar to friction, the line-integral of which is $-ri$, the force being taken per unit of charge.

Thus, if the length of the circumference of the wire be s, the force acting on an electron in the direction of the wire will be

$$-L\frac{e}{s}\frac{di}{dt} - r\frac{e}{s}i.$$

The equation of motion of an electron is thus

$$m\frac{d}{dt}(v + w) = -L\frac{e}{s}\frac{di}{dt} - r\frac{e}{s}i,$$

or

$$\frac{m}{Ne}\frac{di}{dt} + L\frac{e}{s}\frac{di}{dt} + r\frac{e}{s}i = -m\frac{dw}{dt}.$$

If $i = 0$ for $t = t_1$ and $t = t_2$, we find on integration that

$$\frac{re}{s}\int_{t_1}^{t_2} i\,dt = -\Big|mw\Big|_{t_1}^{t_2}.$$

The first transient current is thus $-\frac{mws}{re}$, the second $\frac{mws}{re}$; both depend on the value of e/m. If the moving particles are negative electrons, e/m is great and the effect is accordingly small.

The direction of the current observed by Tolman and Stewart indicated that the particles were negative electrons. In their experiments the wire ring was really a flat, circular coil whose mean diameter was nearly 25 centimeters, the length of the wire being 46,650 centimeters. This coil could be made to rotate about a vertical axis at a speed of 5000 revolutions per minute. The ends of the windings were connected with a galvanometer by means of two wires leading upward to the ceiling, the wires being twisted round each other during the rotation. In this way sufficient freedom of motion was obtained. The coil was brought to rest by means of a brake in a fraction of a second; the period of the galvanometer was about ten seconds.

Many sources of error had to be avoided or compensated, or taken account of in the calculation. I shall mention only one of them. During the rotation the coil expands by centrifugal force and contracts again when brought to rest. This changes the number of lines of magnetic force passing through the coil, the force being the vertical component of the terrestrial magnetic force, and produces induced currents of the same order of magnitude as the currents we want to observe.

Why is a coil used instead of a single wire? If a coil were short-circuited the current would be the same as in a single wire, but a connection with a galvanometer is necessary. The coil must therefore have so many windings that its resistance is not very small in comparison with that of the galvanometer.

The ratio m/e can be derived from the experiments, and, since m_H/e is known, the value of m_H/m can be found, m_H being the mass of an atom of hydrogen. The experimental results were as follows:

1. Length of wire, 46,650 centimeters; eight series of experiments.

$$\frac{m_H}{m} = 1890, 2170, 2120, 1860, 2070, 1790, 1940, 2160.$$

Extremes, 2170 and 1790.

2. Length of wire, 30,370 centimeters; two series of experiments.

$$\frac{m_H}{m} = 1920, 1960.$$

The general mean of these values is 1910, which is not very far from the value 1850 which follows from the best determinations.

21. The Scattering of Light by the Molecules of a Gas. The second case in which the classical theory seems to be all right is that of the scattering of light by a body consisting of molecules, — a phenomenon for which the late Lord Rayleigh found a formula which is universally known and for the discussion of which I should like to make a short digression.

Let us return to the consideration of formulæ (74). They represent the field due to a vibrating electron when we omit the electrostatic part, or, as we may say, they represent the field due to a charge e concentrated in the vibrating particle Q, and a charge $-e$ placed at the position of equilibrium O of Q.

Charges $-e$ at O and e at Q constitute an electric moment of magnitude $e\ OQ$ and direction OQ.

We can therefore say that (74) represents the field of a periodically changing electric moment whose direction is always that of Ox and whose magnitude at time t is

$$ae \cos n(t+p).$$

Formula (78) now shows that a varying electric moment of amplitude s, represented, say, by $s \cos n(t+p)$, gives rise to a radiation of energy amounting to

$$\frac{s^2 n^4}{12\,\pi c^3} \tag{84}$$

per unit of time.

Consider now a beam of light propagated through a medium (a gas, for instance) composed of molecules. In former theories

we supposed the electrons in the molecules to be at rest so long as they were not excited by incident light. Under the influence of rays of light the electrons were supposed to be displaced from their positions of equilibrium. So in each molecule the electric moment would be changing continually with the period of the vibrations of the incident light. The scattering of light in all directions would be precisely the emission by these variable electric moments.

According to modern views the phenomena are less simple. The electrons contained in atoms and molecules are not at rest but are revolving around a nucleus, or a number of nuclei, as the case may be. There are no positions of equilibrium and no quasi-elastic forces by which electrons are driven to such positions. Yet we can imagine that under the electric force of the incident light the motion of the electrons is changed, the electrons being shifted, first in one direction and then in the opposite one, from the positions which they would have if there were no incident light. We may say that the deviation from the original motion constitutes a variable electric moment, and we can imagine that this moment, or rather its changes, produces radiation according to the above formula.

Consider a beam of light going in the direction Ox. At a point of the beam the electric force is in the direction of Oy and is represented by

$$E_y = a \cos n(t + q),$$

say, where a, n, and q are constants. We may say that on an average the electric moment of a particle has the direction of E_y and is proportional to it, being represented by

$$ka \cos n(t + q),$$

where k is a constant depending on the nature of the molecules.

Replacing s by ka in (84), we find that the radiation from one molecule is given by

$$\frac{k^2a^2n^4}{12\,\pi c^3}.$$

We shall now use this formula to calculate the extent to which a beam is weakened by this radiation.

Let the beam have a cross-section of unit area and an intensity I measured by the energy transmitted through this area per unit of time. The difference between the values of I at two sections x and $x + dx$ will be

$$-\frac{dI}{dx}\,dx.$$

This expression for the diminution of I must be equal to the energy radiated by all the molecules between the sections x and $x + dx$. Let the number of molecules per unit of volume be N, then the number between the sections just named is $N\,dx$. Therefore, in order to find the total radiation from all these molecules, we multiply the radiation from a single molecule by $N\,dx$, — a step which will be justified presently. The resulting equation is

$$-\frac{dI}{dx} = \frac{k^2a^2n^4}{12\,\pi c^3}N. \tag{85}$$

Now there is a simple relation between I and a, so that on the right-hand side we get I itself multiplied by a constant. To find this relation we shall suppose the expression on the right-hand side to be very small, which is true in the case of a gas of small density. Then in evaluating a we can proceed as if we had to do with ether without any molecules. In this case the amplitude of the magnetic vibrations would be equal to that of the electric vibrations; in other words, we should have $E_y = H_z$, and Poynting's vector would be

$$ca^2 \cos^2 n(t + q)$$

with a mean value

$$I = \tfrac{1}{2}\,ca^2.$$

Therefore

$$a^2 = \frac{2\,I}{c}.$$

Equation (85) may now be written

$$\frac{dI}{dx} = -\,\alpha I,$$

where

$$\alpha = \frac{k^2n^4N}{6\,\pi c^4}. \tag{86}$$

The solution of this equation is

$$I = I_0\, e^{-\alpha x}, \tag{87}$$

so that α is seen to be the extinction-coefficient.

In the formula for α there is a constant k of which we can say nothing except that it is related to the velocity of propagation and the index of refraction μ.

We wrote kE_y for the electric moment produced in a molecule by the electric force E_y. Thus the electric moment of the system of molecules per unit of volume is

$$kNE_y.$$

This is electric polarization; the dielectric displacement is

$$D_y = E_y + kNE_y = (1 + kN)E_y = \epsilon E_y,$$

and so the dielectric constant is

$$\epsilon = 1 + kN.$$

Hence, by Maxwell's law,

$$\mu^2 = 1 + kN,$$

and so the formula for the coefficient of extinction becomes

$$\alpha = \frac{(\mu^2 - 1)^2 n^4}{6\pi N c^4}, \tag{88}$$

or, since $\quad n = \dfrac{2\pi c}{\lambda}, \quad \dfrac{n^4}{c^4} = \dfrac{16\pi^4}{\lambda^4},$

where λ is the wave-length in ether,

$$\alpha = \frac{8}{3}\frac{\pi^3(\mu^2 - 1)^2}{N\lambda^4} = \frac{32}{3}\frac{\pi^3(\mu - 1)^2}{N\lambda^4}, \tag{89}$$

because $\mu - 1$ is very small.

The above discussion calls for some further considerations.

1. In order to find the total scattering we have simply multiplied the amount of energy radiated from one particle by $N\,dx$, the number of molecules in the layer. One might doubt the correctness of this. Two molecules quite near each other (at a distance apart $<< \lambda$) would acquire electric moments having approximately the same phase and would jointly produce radiation having an intensity four times that of the radiation from one molecule. Yet our calculation is correct on account of the irregular distribution of molecules.

This may be seen from a consideration of the following general theorem: Let a great number of vibrations of equal amplitude but of very different phases be compounded with each other. Assuming for simplicity that the vibrations are all in one direction, we find the resultant vibration to be

$$\sum a \cos n(t+q).$$

The intensity is measured by the square. If we take

$$\sum a^2 \cos^2 n(t+q),$$

or, in the mean, $\Sigma \frac{1}{2} a^2$, we simply get the sum of the individual intensities, but we have to add

$$2 \sum \sum a^2 \cos n(t+q) \cos n(t+q'),$$

the mean value of which is

$$\sum a^2 \cos n(q-q').$$

This will vanish when $\cos n(q-q')$ has as many negative values as positive ones. Perhaps this will be true in one experiment; perhaps it will be true if the experiment is repeated a great number of times and a mean result is taken.

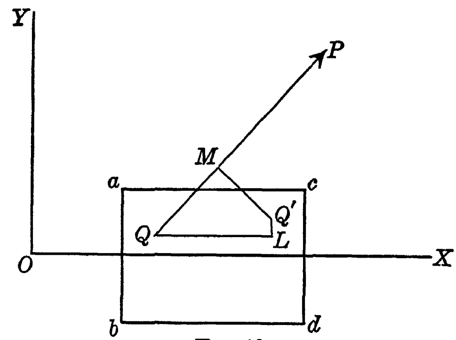

FIG. 19

Now consider a portion *abcd* (Fig. 19) of the light-beam whose length is very great in comparison with the wave-length. In this space we shall place the molecules quite at random (say with closed eyes), so that the probability that any one molecule will lie within a given element of volume is the same for all elements. If Q and Q' are positions of two molecules, or rather the projections of their positions on

a plane XOY in which the distant point P is supposed to lie, the phase-difference is given by

$$QL - QM,$$

$Q'L$ being at right angles to Ox and $Q'M$ to QP.

This phase-difference is a linear function of the coördinates. Now when the experiment of placing the two particles in the cylinder is repeated a great number of times the cosine of $\frac{2\pi}{\lambda}(QL - QM)$ can be shown to have the mean value zero.

2. Thus our result depends on the irregular distribution of molecules which exists in the case of gases. If the molecules were regularly spaced, as in crystals, there would be no scattering. This is true so long as the wave-length of the light is very large compared with the distance between two consecutive molecules. The scattering of light which is sometimes observed in crystals is due to irregularity of structure or to the presence of foreign particles. With rays of very small wave-length there may be a true scattering, which can be limited to definite directions. We then have the phenomenon of Von Laue and the Braggs.

3. The theory can be presented in a quite different form. Let us divide space into equal elements whose volume is small compared with the wave-length λ but is still large enough to contain many molecules. On account of the irregular distribution there will not be exactly the same number of molecules in each element. The fluctuations in density can be calculated by means of the theory of probability.* If n denotes the mean number of molecules in an element, the probable deviation from the mean is of the order of magnitude $\sqrt{n}$. Now differences of density imply differences in the index of refraction (that is, the medium is not optically homogeneous), and this leads to scattering, just as a solution of sugar into which some water has been poured is opaque so long as the mixture has not become homogeneous.

* M. von Smoluchowski, Boltzmann-Festschrift (1904), p. 626; *Ann. d. Phys.*, Bd. 25 (1908), p. 205; *Phil. Mag.*, Vol. 23 (1912), p. 165; *Phys. Zeitschr.*, Bd. 13 (1912), p. 1.

By following this line of thought Rayleigh's formula can again be deduced, the formula being exactly the same as before.* After all, the two methods are substantially the same.

4. The second method is of wider applicability. The deviation from the really homogeneous state can be determined by means of the principles of statistical mechanics, not only for a gas but also for liquid bodies.† Take, for example, a mixture of two substances, either gaseous or liquid. There can be deviations of different kinds from a state of equilibrium. There may be fluctuations in density (such as those we have considered), in temperature, and in composition. All these contribute to the scattering of the rays. Near the critical state (when two phases which can coexist are very slightly different from each other) the fluctuations become larger than usual, and we get the phenomenon of critical opalescence.‡

5. According to the above theory scattered light ought to be plane-polarized when the incident light is so, and more or less partially polarized when the incident light is not polarized.

In the first case, if the incident beam has its electric vibrations along OY, the alternating electric moments of the molecules will be in that same direction, and they will give rise, in any direction OP, to a polarized radiation.

On the other hand, when the incident light is wholly unpolarized, we can decompose the incident vibrations into two com-

* The formula obtained by Smoluchowski is really more general than Rayleigh's formula, since it contains the isothermal compressibility of the substance as a factor. It reduces to Rayleigh's formula in the case of an ideal gas obeying Boyle's law. The general formula has also been obtained by Einstein (*Ann. d. Phys.*, Bd. 33 (1910), p. 1275) by a calculation of the electromagnetic radiation.

† It has been pointed out by C. V. Raman that the observed scattering power of liquids under ordinary conditions agrees with that calculated from the Einstein-Smoluchowski formula. Raman has also expressed the opinion that this formula may determine the magnitude of light-scattering in all fluid media. The formula has recently been tested experimentally by K. R. Ramanathan (*Proc. Roy. Soc. of London*, Ser. A, Vol. 102 (1922), p. 151) for the scattering of light by ether, in the vapor and liquid phases at different temperatures, and has been found to be applicable except in the *immediate neighborhood* of the critical point. This agrees with the result obtained by Keesom for the case of ethylene (*Ann. d. Phys.*, Bd. 35 (1911), p. 591).—Ed.

‡ See also Lorentz, Les théories statistiques en thermodynamique. Leipzig, 1916. References to experimental work on the subject are given in Ramanathan's paper.

ponents along OY and OZ (Fig. 20); these components have in the mean equal intensities and are incoherent in phase.

At P in the scattered light there are now two components in the planes POY and POZ respectively, both perpendicular to OP and having intensities proportional to the squares of the sines of the angles POY and POZ. The phases of these components will be incoherent, and so there will in general be a partial polarization depending on the difference of the sines of the said angles. The polarization will be complete only in the directions OY and OZ.

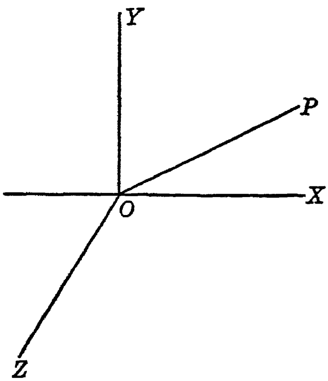

FIG. 20

It may be, however, and it has actually been observed, that the polarization is less than would follow from these considerations.

In order to explain this we can assume that the molecules are anisotropic. For each of three mutually perpendicular principal axes an electric force along the axis produces an electric moment in the same direction and proportional to the force, the three coefficients being k_1, k_2, k_3. We suppose the particles to be orientated in all possible directions. Then the index of refraction μ is found to depend on the mean value

$$\tfrac{1}{3}(k_1 + k_2 + k_3),$$

just as it depends on k in the former case. The calculation of the scattered energy becomes somewhat more complicated, though the method remains the same. It is found that the polarization is less than before. The extinction-coefficient is somewhat different from that of Rayleigh; the former value has to be multiplied by

$$1 + \frac{(k_2 - k_3)^2 + (k_3 - k_1)^2 + (k_1 - k_2)^2}{(k_1 + k_2 + k_3)^2},$$

but this will be of no importance.

Rayleigh's formula has been verified in the first place by astronomical observations; namely, by measuring the intensity of light from the stars seen at different altitudes, so that the length of the path in the air is different in different cases. It has been found that on the top of a mountain, where the air is sufficiently free from dust (for instance, Mt. Wilson), the diminution of intensity due to the passage of the rays through the air may really be accounted for by means of the formula, taking for N the known value. Conversely, by means of the equation we can calculate Avogadro's constant N. Thus we find values in good agreement with those given by other methods.*

To the present Lord Rayleigh we are indebted for most interesting experiments on the subject, by which the theory is also confirmed.†

22. Further Discussion of Rayleigh's Formula. The quantity

$$\alpha = \frac{32\,\pi^3(\mu-1)^2}{3\,N\lambda^4}$$

seems at first sight to be inversely proportional to N. Indeed, if you produce a definite value of μ by a greater or less number of molecules, as you can imagine, α will increase when N decreases; in other words, α increases with the coarse-grainedness. But for a given gas

$$\mu - 1 = \omega N, \tag{90}$$

where ω is a characteristic constant; we then have

$$\alpha = \frac{32\,\pi^3\omega^2 N}{3\,\lambda^4},$$

proportional to N.

23. The Application of Rayleigh's Formula to Atmospheric Problems. Let us suppose that N is a function of x; then we have

$$dI = -\,\alpha I\,dx,$$

* See, for instance, A. Schuster, *Nature*, Vol. 81 (1909), p. 97; L. V. King, *Phil. Trans.*, Ser. A, Vol. 212 (1912), p. 375; *Trans. Roy. Soc. of Canada*, Vol. 9 (1915), p. 99.

† *Proc. Roy. Soc. of London*, Ser. A, Vol. 94, pp. 260, 453; Vol. 95, pp. 155, 476; Vol. 97, p. 435; Vol. 98, p. 57.

where α is a function of x. The solution of this equation is

$$I = I_0 e^{-\int \alpha\, dx},$$

where
$$\int \alpha\, dx = \frac{32\, \pi^3 \omega^2}{3\, \lambda^4} \int N\, dx. \tag{91}$$

The ratio in which the intensity is diminished when light travels over a certain distance is thus seen to depend on the integral $\int N\, dx$, which may be interpreted as the number of molecules contained in a cylinder of cross-section unity placed along the beam of light.

We shall apply our result to the scattering of light in a planetary or stellar atmosphere, supposing the height to which it is practically limited to be very small in comparison with the radius OA of the spherical body. In this case the density of gas may be supposed to vary with the height z (Fig. 21) according to a law of type

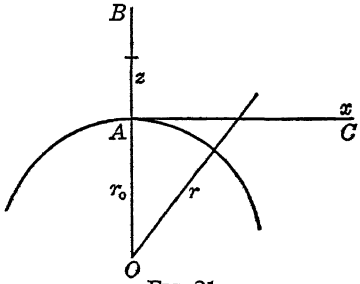

FIG. 21

$$N = N_0 e^{-sz},$$

where s is a constant proportional to the acceleration of gravity and the molecular weight, and inversely proportional to the temperature.

Let us calculate the extinction, for the two extreme cases, of a ray along the vertical BA and of another, CA, tangent to the surface.

Along AB
$$\int N\, dz = \int_0^\infty N_0 e^{-sz}\, dz = \frac{N_0}{s}.$$

Along AC
$$\int N\, dx = \int_0^\infty N_0 e^{-s(r - r_0)}\, dx.$$

Since only the part of AC for which $r - r_0$ is very small in comparison with r_0 contributes appreciably to the integral, we may write

$$\int N\, dx = \int_0^\infty N_0 e^{-\frac{sx^2}{2 r_0}}\, dx = \frac{1}{2} N_0 \sqrt{\frac{2\, \pi r_0}{s}}. \tag{92}$$

In the first case we may also write for the integral $N_0 h$, h being the height of the homogeneous atmosphere. Thus, by (90) and (91), if the index 0 refers to the surface of the earth,

$$\int \alpha\, dx = \frac{32\,\pi^3(\mu_0 - 1)^2 h}{3\,\lambda^4 N_0}.$$

Take $$h = 76\,\frac{13.59}{0.001293} = 7.99 \times 10^5 \text{ cm.},$$

$$\mu_0 = 1.000294, \quad \lambda = 5.9 \times 10^{-5}, \quad N_0 = 2.70 \times 10^{19},$$

then $$\int \alpha\, dx = 0.069, \quad \frac{I}{I_0} = 0.933.$$

This is the amount of extinction over a distance of 8 kilometers in air at ordinary pressure. The result can of course be applied to a beam traveling in a horizontal direction. At a distance of 260 kilometers the transmitted light would be no more than 0.1 of the incident.

When we pass to air of another density, we find that the length of path required for a given extinction will be inversely proportional to the density. Thus, if the pressure were 0.01 atmosphere the intensity would be reduced to 0.1 of its original value along a path of 26,000 kilometers.

Considerations like these must be kept in mind in the discussion of certain solar phenomena.

Taking into account the rates of diminution of density in a vertical direction, and the refractive index, we can easily calculate that in our atmosphere a ray of light traveling at a low altitude in a direction nearly parallel to the earth's surface will be bent downward to such a degree that its radius of curvature is about 4.3 times the radius of the earth. It has been suggested by Schmidt that the sun consists of a mass of gas whose density increases so rapidly toward the center that the radius of curvature of a ray of light is even less than the so-called radius of the sun. Schmidt also maintains that the sharp boundary is an optical illusion depending on the fact that some of the rays that start from a point A are bent so much that they do not get outside a certain critical sphere S. A ray AE which reaches the eye

may describe a curved path which is for some distance nearly parallel to the surface S (Fig. 22). I shall not go into these questions, and I shall only say that in speaking of them one must not forget that, on account of the scattering, a ray of light can hardly be expected to travel over paths of the length shown in the diagram. The sun's radius is 6.9×10^5 kilometers and therefore the above-mentioned length of 26,000 kilometers is no more than $\frac{1}{166}$ of the circumference; it will subtend an angle of not much over 2° at the center.

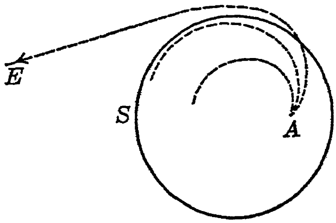

FIG. 22

Returning to the earth, we may remark that if the height of the homogeneous atmosphere were twenty times as great as it actually is, the intensity of the light transmitted from a star in the zenith would be only 0.25 of the intensity of the incident light, and if the height were forty times as great as it actually is, the ratio of intensities would be only 0.06. It would then seem as if we were living in a dense fog. However, and this is a curious theoretical result, just as in the case of a fog, objects would be seen with sharp boundaries.

The phenomena that would be observed in a medium in which rays of light were continually scattered in all directions were examined theoretically by Schuster many years ago.* They become rather complicated when repeated scattering is taken into account; we shall therefore consider only one simplified problem. Let us suppose that above the perfectly black plate AB (Fig. 23) there

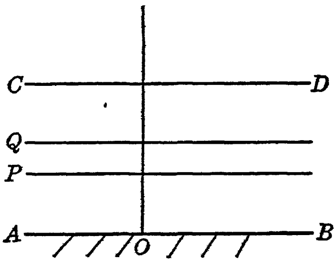

FIG. 23

* A. Schuster, "Radiation through a Foggy Atmosphere," *Astrophys. Journ.*, Vol. 21 (1905), p. 1. See also W. H. Jackson, *Bull. Amer. Math. Soc.*, Vol. 16 (1910), p. 473; L. V. King, *Phil. Trans.*, Vol. 212 (1912), p. 375. K. Schwarzschild, *Berlin. Sitzungsber.* (1914), p. 1183; E. A. Milne, *Phil. Trans.*, Vol. 223 (1922), p. 21.

is a layer extending to CD, in which the properties are everywhere the same. Let us divide the layer into infinitely thin layers of thickness dx.

The black body radiates and the rays are thrown to and fro by the different layers. All rays that return to AB are absorbed.

We shall fix our attention on the light scattered by all the particles contained in a layer. The black radiation coming from AB is distributed over different directions in a definite manner and is for each direction unpolarized, and we shall now suppose that the rays coming from any thin layer have just the same distribution and are equally unpolarized. With these suppositions the aggregate of rays passing through a plane parallel to AB either in an upward or in a downward direction may be characterized by the total flux of energy per unit surface and unit time.

At a plane P whose distance from AB is x we shall suppose the upward flow to have the intensity i and the downward flow the intensity i', the quantities i and i' being functions of x.

At the plane Q the corresponding quantities are supposed to be $x + dx$, $i + di$, and $i' + di'$ respectively.

Let us suppose that light of intensity i, falling on the layer PQ of thickness dx, produces scattered light of intensity $qi\,dx$ coming back and light of intensity $i(1 - q\,dx)$ passing through. The coefficient q will be constant all through the space between AB and CD, and we have the equations

$$i + di = (1 - q\,dx)i + q\,dx(i' + di'),$$
$$i' = (1 - q\,dx)(i' + di') + q\,dx \cdot i.$$

Omitting products of small quantities, such as dx and di', we have

$$di = q(i' - i)\,dx, \quad di' = q(i' - i)\,dx,$$
$$di' = di, \qquad i' = i - C,$$
$$\frac{di}{dx} = -\,Cq, \quad i = -\,Cqx + i_0, \quad i' = -\,Cqx - C + i_0.$$

Here i_0 is the flow of energy at AB due to the radiation of the black body. To determine the constant C we suppose that

$i' = 0$ at the upper boundary of the layer where $x = l$, say. This will be true when there is no appreciable reflection at the boundary CD. Thus

$$C = \frac{i_0}{1 + ql},$$

$$i = i_0 \frac{1 + q(l - x)}{1 + ql}.$$

At the upper boundary $$i = i_0(1 + ql)^{-1}. \tag{93}$$

This is the intensity of the light which emerges from the layer. When $l \rightarrow \infty$, $i \rightarrow 0$. If we neglected scattering, we should have

$$di = -qi\,dx,$$

$$i = i_0 e^{-qx},$$

and, for $x = l$, $$i = i_0 e^{-ql},$$

which is less than (93).

24. The Sun's Apparent Boundary. C. G. Abbot, in his book on the sun, has given a theory of the sun's disk which depends on the scattering of light by the particles of the solar atmosphere. Such a theory may, I think, be developed as follows:

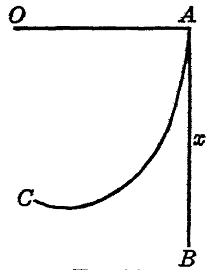

FIG. 24

On account of the scattering we cannot look very deep into the sun, or, rather, the light that reaches us cannot come from any considerable depth. For the sake of definiteness we shall say that the light coming from a point in the globe becomes imperceptible when its intensity is diminished to 1 per cent of the original value. This will enable us to learn, indirectly at least, something about what we may call the impenetrable layer at the sun's surface. We shall consider the part of this layer near the border of the sun, and we shall show that the pressure at a point just inside the layer can be calculated.

Let O be the sun's center. We want to find a point A such that, when AB, at right angles to OA (Fig. 24), is directed toward the observer,

$$e^{-\int \alpha\,dx} = 0.01,$$

where the integral extends along AB to infinity. Indeed, this means that a ray going from A to B will have in the end no more than 1 per cent of its initial energy. For this it is necessary that

$$\int \alpha\, dx = 4.60. \tag{94}$$

Let $OA = r$ and let N be the number of molecules per unit volume at altitude z above the sphere AC; then the law of variation of density is

$$N = N_0 e^{-sz}, \tag{95}$$

where $N = N_0$ on the sphere AC. The integral in (94) is given by (91); thus

$$\frac{32\,\pi^3\omega^2}{3\,\lambda^4}\int N\, dx = 4.60,$$

or, taking into account (92),

$$\frac{32\,\pi^3\omega^2}{3\,\lambda^4}\,\frac{1}{2}\,N_0\sqrt{\frac{2\,\pi r_0}{s}} = 4.60,$$

or

$$N_0 = 4.60\,\frac{3\,\lambda^4}{16\,\pi^3\omega^2}\sqrt{\frac{s}{2\,\pi r_0}}.$$

Let us suppose that the solar atmosphere in the neighborhood of A consists of hydrogen at a temperature of 6000° absolute. Let g be the acceleration of gravity at the sun's surface and ρ the density of hydrogen. Then the pressure is $p = \beta\rho$ where β is a constant. Furthermore, since

$$\frac{dp}{dz} = -g\rho,$$

we have

$$\frac{d\rho}{dz} = -\frac{g}{\beta}\rho,$$

and so

$$s = \frac{g}{\beta}.$$

For hydrogen at 273° abs.

$$\frac{p}{\rho} = \frac{1.0132\cdot 10^6}{0.0899\cdot 10^{-3}},$$

and at 6000°

$$\beta = \frac{1.0132\cdot 10^6\cdot 6000}{0.0899\cdot 10^{-3}\cdot 273},$$

$$\log\beta = 11.3939;$$

and since the acceleration of gravity on the sun is 27.6 times that on the earth,

$$\log s = 3.0389 - 10.$$

For hydrogen at 0° and a pressure of 76 centimeters the value of μ for the D line of sodium is 1.0001384; also

$$\omega = \frac{\mu - 1}{N}, \quad \text{where} \quad N = 2.70 \cdot 10^{19}.$$

$$\log \omega = 6.7098 - 30.$$

We may use this value if we suppose μ to depend only on the density and not on the temperature.

Again, $r_0 = 6.95 \cdot 10^{10}$ cm.

and $\lambda = 5.89 \cdot 10^{-5}$ for the D line.

The result is that

$$\log N_0 = 18.8047, \quad N_0 = 6.4 \cdot 10^{18},$$

which corresponds to a pressure of 5.2 atmospheres.

We could also assume the hydrogen to be dissociated, so that we have to reckon with atoms. Then, if we suppose that this dissociation, if it takes place without change of density, does not affect the refractive index, ω becomes half as great. The coefficient s changes in the same ratio, so that N_0 and the pressure corresponding to it are multiplied by $2^{\frac{3}{2}}$. Our result thus becomes nearly 15 atmospheres.

If, instead of requiring a reduction of intensity to 0.01, we were satisfied with a reduction to 0.1, we should have to replace 4.60 by 2.30, so that in the above cases N_0 would become half as large as before and correspond to pressures of 2.6 and 7.5 atmospheres.

We have now found something about the depth of the impenetrable layer at the limb of the sun, we may say of the layer that can be directly seen. Some of the light coming from beneath it may be seen by scattering, but we shall not speak of this and shall only put the question, Will this layer itself be seen with a sharp boundary? This depends entirely on the way in which the luminosity falls off when we go outward from

the point considered. It is very difficult to give a definite answer, since we ought to know more about the state of things in the solar atmosphere.

We can, however, calculate the height above A at which the density of the gas would be, for instance, $\frac{1}{100}$ of its value at A. This height z, at which we may suppose the observed luminosity to be much smaller than that which is seen at A, is determined by

$$e^{-sz} = 0.01,$$

or

$$sz = 4.60, \quad z = \frac{4.60}{s},$$

and with the above value of s

$$z = 4.2 \cdot 10^7 \text{ cm.},$$

which is about the 1600th part of the radius of the sun.

There would then be a rather sharp boundary. If we were to reckon with atoms, we should find just double the above value, that is, $\frac{1}{800}$ of the radius.

25. The Scattering of Röntgen-Rays. The scattering of Röntgen-rays can be treated in the same way as that of light. Let us suppose again that at a point of the incident beam

$$E_y = a \cos n(t + q).$$

We have considered the electric moment of one particle, representing its amplitude by ka and determining k in terms of the index of refraction. We shall now fix our attention on one electron and understand by η its displacement in the direction of the y-axis; then the equation of motion is

$$m\ddot{\eta} = ea \cos n(t + q) + Y,$$

where Y is the force which, in addition to eE_y, may act on the electron. If Y is neglected, we get the particular integral

$$\eta = -\frac{ea}{mn^2} \cos n(t + q).$$

When can Y be neglected? Let us suppose that Y is a quasi-elastic force $-k\eta$; then the particular integral is

$$\eta = \frac{ea}{k - mn^2} \cos n(t + q),$$

and the suppression of Y is justifiable if $k << mn^2$, that is, if $n_0 << n$, where

$$n_0 = \sqrt{\frac{k}{m}}$$

is the frequency of the vibrations which the electron could perform under the influence of Y alone.

In general Y can be neglected if the electron is bound loosely enough. That it may be assumed to be so in the case of Röntgen-rays is due to the very high frequency of these rays.

Now if the electron has a simple periodic motion, we can find the radiation of energy by multiplying the square of the amplitude by

$$\frac{e^2n^4}{12\,\pi c^3};$$

since the amplitude is $\frac{ea}{mn^2}$, we find

$$\frac{e^2n^4}{12\,\pi c^3}\cdot\frac{e^2a^2}{m^2n^4}=\frac{e^4a^2}{12\,\pi m^2c^3}.$$

If there are N electrons in the molecule, the secondary radiation from the molecule will be

$$\frac{Na^2e^4}{12\,\pi m^2c^3}. \tag{96}$$

This is equivalent to the formula found many years ago by Sir J. J. Thomson * and applied by Barkla to his experiments on the production of secondary Röntgen-rays by the scattering of primary ones. By means of it he was able to obtain an estimate of N, the number of electrons in a molecule or atom. You know that this was the first determination of the kind.

One or two remarks may be made. The frequency n has disappeared when the final formula is obtained, and it does not occur in the expression for the flow of energy in the incident beam, which is simply $\frac{1}{2}\,a^2$. We can therefore deduce N from a comparison of the intensities of the primary and secondary rays without knowing the frequency or wave length of the

* Conduction of Electricity through Gases, 2d ed., p. 325.

rays. This is a fortunate circumstance, because at the time of Barkla's experiments the frequency of X-rays was unknown.

Of course, Rayleigh's formula

$$\alpha = \frac{32\,\pi^3(\mu - 1)^2}{3\,\lambda^4 N},$$

also applies to X-rays. From it we may conclude that in comparison with light μ must be little different from 1, because λ^4 is so small that this cannot be compensated by the observed value of the index of extinction α. In fact μ differs so little from unity that the refraction of X-rays has for a long time escaped observation.

26. The Propagation of Light through a Ponderable Body. We have now discussed some phenomena which seem to prove that vibrating electrons really radiate energy. If we recognize this, we are encouraged to use this same idea for other purposes. Thus we can deduce the laws of propagation of waves of light in ponderable bodies by simply considering the vibrations that are produced in these molecules and the radiation from the molecules to which these vibrations give rise. The waves thus produced are superposed on the incident waves and interfere with them. By studying the effect of this interference (of which we did not speak when we considered the scattering of light) we find the influence which ponderable matter has on the velocity of propagation. This may be worked out for a medium of any density, but it will suffice here to consider the propagation in a gas of small density, such that the influence on the propagation can be treated as infinitely small. Yet the number of molecules, even in a layer of a thickness small in comparison with λ, will be taken to be very large. Consider a beam of light passing in the direction Ox and suppose the gas to be bounded by two glass plates CD and EF (Fig. 25) at right angles to Ox. In the

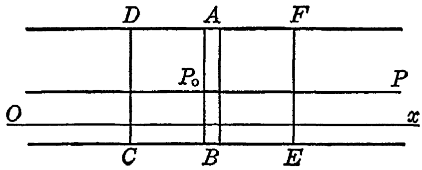

FIG. 25

outside space there is only ether, and we want to determine the vibrations that are received by some distant point P, for whose coördinates we shall write x', y', z'. We shall write

$$E_y = a \cos n\left(t - \frac{x}{c} + p\right) \tag{97}$$

for the electric force in the primary beam, the amplitude a being constant over the greater part of the cross-section AB and diminishing to zero at its edges. The electric force at P will consist of two parts, of which the first,

$$E'_y = a \cos n\left(t - \frac{x'}{c} + p\right), \tag{98}$$

is due to the primary beam, and the second, E''_y, to the vibrations that are emitted by the molecules between the planes CD and EF. Now under the influence of the electric force (97) a molecule of the air situated at the point (x, y, z) will have an electric moment

$$p_y = ka \cos n\left(t - \frac{x}{c} + p\right),$$

and by this it will contribute to E''_y at the point P a part, for which we find by our former formulæ, writing r for the distance of the point (x, y, z) from P, the expression

$$\frac{kan^2}{4\pi c^2}\left[1 - \frac{(y'-y)^2}{r^2}\right]\frac{1}{r}\cos n\left(t - \frac{x+r}{c} + p\right).$$

Now all the molecules situated in an infinitely thin transverse layer may just as well be taken to lie in the plane AB. Let A be the number of these per unit of area; then we have to calculate

$$\frac{kn^2 A}{4\pi c^2}\int\left[1 - \frac{(y'-y)^2}{r^2}\right]\frac{a}{r}\cos n\left(t - \frac{x+r}{c} + p\right)d\sigma$$

where $d\sigma$ is an element of the plane AB. We shall write the integral in the form

$$I = \int \Omega \cos n\left(t - \frac{x+r}{c} + p\right)d\sigma.$$

As a point moves across the plane AB, Ω varies slowly, while the cosine factor varies rapidly. We can divide the plane AB into a central part and surrounding zones, so that the cosine has different signs in adjacent zones. These annular regions are the zones of Fresnel, first used in his theory of diffraction. The boundary lines between the zones are circles (full circles or parts of them, as the case may be) around the projection P_0 of P on AB; the distances of these circles from P form an arithmetical progression with difference $\frac{1}{2}\lambda$.

These zones are very numerous; consequently their contributions to the integral nearly destroy each other. There is, however, a small quantity left which we desire to know and which can be calculated by a beautiful method due to Kirchhoff.*

Let r_0 be the smallest value of r; that is, the distance $P_0P = x' - x$. Let r_1 be the greatest value (corresponding to a point somewhere on the boundary of AB) and let r be an intermediate value. Let us put

$$\int_{r_0}^{r} \Omega\, d\sigma = F(r), \tag{99}$$

the integral being extended to the part of the section AB where the distance from P lies between r_0 and r. Then the integral

$$\int \Omega\, d\sigma,$$

taken over the annular element between r and $r + dr$, will be

$$\frac{dF(r)}{dr}\, dr.$$

For all points in that annular element

$$\cos n\left(t - \frac{x+r}{c} + p\right)$$

has the same value; therefore

$$\frac{dF(r)}{dr} \cos n\left(t - \frac{x+r}{c} + p\right) dr$$

* "Zur Theorie der Lichtstrahlen," *Ann. d. Phys. u. Chem.*, Bd. 18 (1883), p. 663; *Berlin. Sitzungsber.* (1882), p. 641.

will be the part of I that corresponds to the annular element determined by the interval dr. Thus

$$I = \int_{r_0}^{r_1} \frac{dF(r)}{dr} \cos n\left(t - \frac{x+r}{c} + p\right) dr. \tag{100}$$

Now Ω, $F(r)$, and $\frac{dF(r)}{dr}$ are slowly variable quantities, while the cosine is a rapidly varying quantity. Therefore in calculating the definite integral we may proceed as if $\frac{dF(r)}{dr}$ were constant for a certain range of r. A more accurate method is to integrate by parts, obtaining the equation

$$I = -\frac{c}{n}\left|\frac{dF(r)}{dr} \sin n\left(t - \frac{x+r}{c} + p\right)\right|_{r_0}^{r_1}$$
$$+ \frac{c}{n}\int_{r_0}^{r_1} \frac{d^2F(r)}{dr^2} \sin n\left(t - \frac{x+r}{c} + p\right) dr. \tag{101}$$

The last term (on account of the factor n in the denominator) is very small in comparison with I itself; we shall therefore drop it. The first term becomes zero at the upper limit $r = r_1$, because at the boundary $a = 0$ and therefore $\Omega = 0$. Thus, finally,

$$I = \frac{c}{n}\left[\frac{dF(r)}{dr}\right]_{r=r_0} \sin n\left(t - \frac{x+r_0}{c} + p\right). \tag{102}$$

We want to find $\frac{dF(r)}{dr}$ for $r = r_0$, but we must not substitute $r = r_0$ in $F(r)$ before differentiation. It is sufficient, however, to know the value of $F(r)$ for values of r very little different from r_0; that is, over the portion of AB in the immediate neighborhood of P_0.

Now $F(r)$ was defined by (99), and, omitting for a moment the factor

$$\frac{kn^2A}{4\pi c^2},$$

we have

$$\Omega = \left[1 - \frac{(y'-y)^2}{r^2}\right]\frac{a}{r}.$$

Moreover, in a small region around P_0, a is constant.

Again, it is clear that if, instead of $(y'-y)^2$, we had $(z'-z)^2$, the result would be exactly the same. Hence we may replace $(y'-y)^2$ by

$$\tfrac{1}{2}[(y'-y)^2+(z'-z)^2]=\tfrac{1}{2}[r^2-(x'-x)^2]=\tfrac{1}{2}(r^2-r_0^2),$$

and Ω by
$$\frac{a}{2}\left(\frac{1}{r}+\frac{r_0^2}{r^3}\right).$$

Let us now take for $d\sigma$ an annular element between circles around P_0 with radii ρ and $\rho+d\rho$ respectively. Then, since $\rho^2=r^2-r_0^2$,

$$d\sigma=2\,\pi\rho\,d\rho=2\,\pi r\,dr,$$

$$F(r)=\int_{r_0}^{r}2\,\pi r\,dr\cdot\frac{a}{2}\left(\frac{1}{r}+\frac{r_0^2}{r^3}\right)=\pi a(r-r_0)-\pi a r_0^2\left(\frac{1}{r}-\frac{1}{r_0}\right),$$

$$\frac{dF(r)}{dr}=\pi a\left(1+\frac{r_0^2}{r^2}\right),\quad\left[\frac{dF(r)}{dr}\right]_{r=r_0}=2\,\pi a,$$

$$I=\frac{c}{n}\cdot 2\,\pi a\cdot\sin n\left(t-\frac{x+r_0}{c}+p\right).$$

Thus, finally, the contribution to E_y'' at the point P, due to molecules in the plane AB, is

$$\frac{kn^2A}{4\,\pi c^2}\cdot\frac{2\,\pi ac}{n}\cdot\sin n\left(t-\frac{x+r_0}{c}+p\right)$$
$$=\frac{knAa}{2\,c}\sin n\left(t-\frac{x+r_0}{c}+p\right). \qquad (103)$$

You will have observed that this result is due to the value which a certain function had for $r=r_0$, and to this extent it can be said that the vibration produced at P by the particles in the plane AB depends on the state of things at P_0. It is as if each vibration proceeded along a straight line, like P_0P; but whereas the vibration due to a particle at P_0 would have been

$$\frac{kan^2}{4\,\pi c^2}\cdot\frac{1}{r_0}\cos n\left(t-\frac{x+r_0}{c}+p\right),$$

we now have the symbol *sine* instead of *cosine*. This means that the phase at P is one quarter of a period behind that which we should find if we had a simple vibration propagated from P_0 to P.

Indeed, vibrations come not only from P_0 but also from the neighboring points of the plane AB; we can therefore express our result by saying that the vibrations due to the different zones of which we spoke destroy each other for the greater part, but that those that are excited by the molecules in a part of the central zone remain.

This central zone and its effective part increase in area when the distance P_0P increases, and this explains why the resulting amplitude is independent of the distance r_0. We see further, by (103), that the phase of the resulting vibration is likewise independent of the position of the plane AB. We may, in fact, write $x + r_0 = x'$, the x-coördinate of P. Let us now suppose that the gaseous medium extends from C to E over a length l. If all the molecules in the layer dx are considered as lying in the plane AB, we shall have

$$A = N\,dx,$$

where N has its former meaning.

The factor by which dx is multiplied is the same for all layers, and we find, therefore, that

$$E_y'' = \frac{knNla}{2\,c} \sin n\left(t - \frac{x'}{c} + p\right). \tag{104}$$

We must add this to the value of E_y' given by (98).

It is to be understood that all this is correct only when the influence of the gas, and therefore the expression (104), is very small; if this were not the case, we could not suppose the electric moment of all molecules to have the same amplitude ka. Thus (104) is infinitely small in comparison with (98).

When a harmonic vibration is compounded with a second vibration of the same kind but of infinitely small amplitude, and with a difference of phase of a quarter period, this amounts to a small change in phase. Indeed, for the sum of E_y' and E_y'' we can write

$$E_y = a \cos n\left(t - \frac{x'}{c} + p - \frac{kNl}{2\,c}\right). \tag{105}$$

If we find a change of phase on the interposition of a column of some substance in the path of a beam of light, we may conclude

that the velocity of light in the substance is different from that in the ether, and the index of refraction μ different from unity.

On account of the interposition the time required for a transmission over the length l is changed from $\frac{l}{c}$ to $\mu\frac{l}{c}$, so that in

$$a \cos n\left(t-\frac{x'}{c}+p\right)$$

we have to replace t by $t-\frac{(\mu-1)l}{c}$, thus obtaining

$$E_y = a \cos n\left[t-\frac{x'}{c}+p-\frac{(\mu-1)l}{c}\right]. \tag{106}$$

Comparing this with our former formula (105), we obtain

$$\mu - 1 = \tfrac{1}{2}kN, \tag{107}$$

and this agrees with the value which we found for μ^2-1 by another mode of reasoning.

So, after all, you see that we have good reasons for believing that electrons with a non-uniform motion really radiate; they did so in all these cases in which they were first set in vibration by incident light. I might add that the interference phenomena that occur when Röntgen-rays fall on crystals can be accounted for by the same principle. Last year W. L. Bragg, Jones, and Bosanquet* showed that the intensity of the Röntgen-rays that are reflected (or diffracted) by rock salt can be calculated by means of the classical theory.

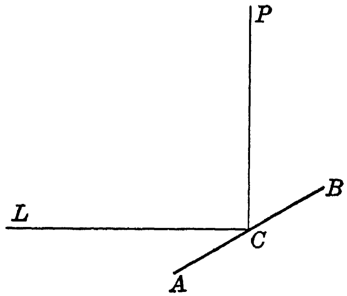

FIG. 26

The method by which we found the effect due to a layer of molecules can often be used to advantage. For instance, Röntgen-rays coming from L are reflected by a layer AB of molecules in a crystal (Fig. 26). The resulting vibration at some point P will be due to a small part of the plane AB quite

* *Phil. Mag.* (6), Vol. 41 (1921), p. 309.

near the point C of regular reflection and will differ somewhat in phase from the vibration that would have traveled over the length of path LCP.*

27. Electromotive Forces. We have found good reasons for believing that electric charges which are not moving uniformly along straight lines can really produce radiation, but we can imagine other causes that may do so as well.

If we were concerned with the propagation of sound in air, the analogue of our moving electron would be a body vibrating in the air. But we can suppose, instead of this, that some external force acts on the air itself; if you could handle air, you would of course be able to set it in vibration. In a similar way, electromotive forces can be supposed to act on electricity. The idea has often been introduced in the case of conducting bodies, but it will serve just as well for dielectrics. You know that in conductors, when there is only what we call electric force (the $\mathbf{E}$ of our formulæ), we have for the current

$$\mathbf{C} = \sigma \mathbf{E}, \tag{108}$$

where σ is the coefficient of conductivity. If there is an external or impressed force acting on electricity (and this is what we shall call electromotive force), the current is given by the vector equation†

$$\mathbf{C} = \sigma(\mathbf{E} + \mathbf{E}_e). \tag{109}$$

Similarly, when we say that in a dielectric there is an electromotive force $\mathbf{E}_e$, we mean that

$$\mathbf{D} = \epsilon \mathbf{E} \tag{110}$$

must be replaced by

$$\mathbf{D} = \epsilon(\mathbf{E} + \mathbf{E}_e). \tag{111}$$

We may even apply this to the ether.‡ Whereas thus far we have not had to distinguish for this medium between $\mathbf{E}$ and $\mathbf{D}$, we now have to do so. We take for the relation between them

$$\mathbf{D} = \mathbf{E} + \mathbf{E}_e, \tag{112}$$

* See Note 5, Appendix.

† Encyklopädie der Mathematischen Wissenschaften, Bd. V, § 14, p. 89.

‡ These ideas are used in a different manner in the theory of radiation by L. Silberstein, *Phil. Mag.* (6), Vol. 29 (1915), p. 709; Vol. 30 (1915), pp. 163, 784. Another application of electromotive forces has been made by the author in *Proc. Nat. Acad. Sci.*, Vol. 8 (1922), p. 333.

which means that dielectric displacement consists of two parts, one due to electric force and the other to electromotive force. It is natural to consider these parts as proportional to the forces by which they are produced. That they are numerically equal to them is due to the choice of units.

On account of the parallelism between electric and magnetic quantities expressed by the fundamental equations we can also introduce the idea of magnetomotive force, as was done long ago by Heaviside. We have only to distinguish (in the ether also) between $\mathbf{H}$ and $\mathbf{B}$ and to put

$$\mathbf{B} = \mathbf{H} + \mathbf{H}_e. \tag{113}$$

If $\mathbf{E}_e$ and $\mathbf{H}_e$ are given as functions of the coördinates and the time, the field can be completely determined. The two problems in which we are given only $\mathbf{E}_e$ and only $\mathbf{H}_e$ respectively are mathematically the same, and the first is very similar to that of finding the field produced by the motion of charges. We can again define a vector and a scalar potential, both of which will be retarded, and which will be determined by formulæ in which, instead of ρ and $\rho\mathbf{v}$, you have quantities depending on $\mathbf{E}_e$. For example, let $\mathbf{E}_e$ be confined to a very small space around the point Q, and let it, in that space, have everywhere the direction of Ox and the magnitude $a \cos n(t + p)$, changing its sign periodically; then in the surrounding field there will be similar vibrations and radiation, as we found in the case of an oscillating electron.* Of course the amount of energy radiated is now equal to the work of the electromotive force, which, per unit of time and unit of volume, is equal to the scalar product $(\mathbf{E}_e \cdot \dot{\mathbf{D}})$.

These electromotive and magnetomotive forces are purely imaginary, but this need not prevent us from introducing them; it will not be the first time that we have done something of the sort. All that can be required is to develop a system of notions, consistent in itself and without internal contradictions, by means of which observed phenomena can be coördinated and accounted for; and if this can be done by

* See Note 6, Appendix.

means of electromotive and magnetomotive forces, for whose directions, magnitudes, and mode of action we can give definite rules, it will be wholly satisfactory.

28. Remarks on the Discrepancy between Old and New Theories. We shall now revert once more to the discrepancy between the old and new theories, of which I have already spoken; that is, to the necessity of supposing that electrons revolving around the nucleus of an atom do not radiate, in spite of the contrary conclusion drawn from Maxwell's equations. If we want to maintain these equations without new terms or other changes, the difficulty cannot be got rid of. It is just the same as if we wanted a body to be moving to and fro in a mass of air or water, or on the surface of water, without producing waves. It is not to be expected that one could make a body oscillate without producing at some point in its neighborhood a state of things that is not continually the same but is changing periodically, and it is in the very nature of things that changes of this kind are propagated outward, carrying a certain amount of energy with them.

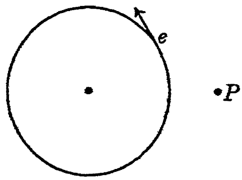

FIG. 27

So also, when an electron *e* describes a closed line, it is clear that a neighboring point *P* will not at all instants be equally affected by the influence of *e* (Fig. 27), and so at *P* there will be periodic changes; Maxwell's equations require them to be propagated.

There is even direct evidence of this periodicity of the field produced by a revolving electron. According to Bohr's theory of spectra in its simplest form certain lines of ionized helium ought to coincide with lines of the hydrogen spectrum. There are, however, small deviations, and it has been found possible to explain them very satisfactorily by taking into account the fact that under the action of the force exerted by the electron the atomic nucleus also describes a small closed curve. In fact, the nucleus and the electron revolve about their common center

of gravity just like the sun and a planet. This is one of the most splendid results of the modern theory of spectra, and I think we have in it actual proof of the motion of the nucleus.

But this motion could never exist at all if the field at O produced by e were not continually changing. But if the field changes at O, why should it not change at P or at any other point, whatever be the distance from O? These periodic changes at distant places are exactly what we mean by radiation.

We may here remark that the radiation becomes considerably less when, instead of one revolving electron, we have two or more, moving exactly or approximately along the same path. Suppose, for instance, that two move in a circle so as to be always diametrically opposite each other. The accelerations to which the radiation is due are equal and opposite, so that if the two particles were at exactly the same place there would be no radiation at all. That there is in reality some radiation is due solely to the fact that there is a certain distance between the electrons; but if this distance is very small in comparison with the wave length, the vibrations imparted to some distant point will be in nearly opposite phases, and the residual radiation will be very feeble.*

It will be still more so if, instead of two electrons, we have three or more of them, equally spaced around a circle. The calculation is to be found in Schott's book on "Electromagnetic Radiation."

Finally, if one had a distribution of charge in the form of a circular ring (the charge per unit length being the same at all points), and if this ring were to rotate with constant angular velocity about its geometrical axis, there would be no radiation at all. The same thing is true in other cases which one can easily imagine. Suppose, for instance, that a charge with uniform volume density moves in a closed tube of any form and variable section, the motion being stationary and similar to that of a stream of incompressible fluid. Then, if we look at the system from some external point, the aspect will always be the same; in other words, nothing changes in the relation of

* See Note 7, Appendix.

the system to some distant point, so that there is no reason for vibrations. This is easily verified by the formulæ. So we might invent distributions and motions of charges that do not radiate; but all endeavors to cope with our difficulty in this way must be considered of no use, since we should lose all the important and beautiful results that have been found by Sommerfeld and Epstein in the further development of Bohr's theory. In the theory of the fine structure of spectral lines, which we owe to the first of these physicists,* and of the decomposition of the lines (Stark effect), which was developed by Dr. Epstein,† it is quite essential that the moving charges should be concentrated in negative electrons with the ordinary e/m. It seems hopeless to try to find the same results with the supposition of some other distribution of charges.

Moreover, the motion of the nucleus, of which I have already spoken, proves that there is around it a revolving electron and not a rotating ring; for of course a ring would not act on the nucleus with a resultant force, and would therefore leave it at rest.

So we are still confronted with the mystery that revolving electrons do not radiate. You will remember that we were driven to our conclusion by our retarded potentials, — the solution, for instance, of the equation

$$\Box \phi = -f(t, x, y, z)$$

being
$$\phi = \frac{1}{4\pi}\int \frac{1}{r} f\left(t - \frac{r}{c}, x, y, z\right) dV.$$

Now it is remarkable that we can just as well satisfy the equation by putting

$$\phi = \frac{1}{4\pi}\int \frac{1}{r} f\left(t + \frac{r}{c}, x, y, z\right) dV,$$

and this would be an accelerated potential. That this is a solution is seen immediately by simply changing the sign of c.

* *Ann. d. Phys.*, Bd. 51 (1916), p. 125; *Münchener Sitzungsber.* (1915), pp. 425–459; Sommerfeld, Atombau und Spektrallinien.

† *Ann. d. Phys.*, Bd. 50 (1916), p. 489; Bd. 51 (1916), p. 168.

If we had used this value of ϕ and the corresponding expressions for the vector potential, the field due to the vibrating electron at a great distance would have become

$$E_x = \frac{aen^2}{4\pi c^2}\left(1 - \frac{x^2}{r^2}\right)\frac{1}{r}\cos n\left(t + \frac{r}{c} + p\right),$$

$$H_y = -\frac{aen^2}{4\pi c^2}\frac{z}{r}\cdot\frac{1}{r}\cos n\left(t + \frac{r}{c} + p\right),$$

$$H_z = \frac{aen^2}{4\pi c^2}\frac{y}{r}\cdot\frac{1}{r}\cos n\left(t + \frac{r}{c} + p\right).$$

The interpretation of these formulæ is that the radiation is toward the vibrating electron. It is not a radiation but an inflow of energy. We can also find a solution of the same kind for the case of an electron having an arbitrary motion of translation. Then if, at a certain moment, the electron has an acceleration, it will be the center of a contracting wave moving toward it.

Now if we have two solutions of Maxwell's equations, one with an efflux and the other with an inflow of energy, we can expect that by properly compounding them we shall find a solution with neither an efflux nor an inflow, but, at the most, certain fluctuations of energy.* In the case of vibrations we shall have standing waves.

But, though this is a correct mathematical solution, I think no physicist will be satisfied with it. It would be absolutely different from the way of thinking about natural phenomena to which we have become accustomed, and not without good reason. We have become familiar with the idea that a change which takes place at a definite place and time produces an effect that spreads outward. It seems difficult indeed to imagine that an experiment which you are going to make, say, tomorrow, in which an electron is to have its motion changed, was prepared in nature many years ago, for if, at the moment at which you

* The matter has been fully investigated by Leigh Page, *Phys. Review*, Vol. 24 (1924), p. 296, and by Nordström, *Proc. Acad. Amsterdam*, Vol. 22 (1920), p. 145. — ED.

produce the acceleration, a wave contracts just at the place of the electron, that wave must have existed long ago at a distance.

Difficulties of the same kind arise in the hypothesis of standing waves. Let me only say that if we find difficulty in imagining waves rushing from all sides to a movable electron, neither can we imagine the standing waves of which that inward propagation forms an essential part. But, after all, we must try to master the difficulty of the absence of radiation. Let me therefore propose to you a solution which, though it does not pretend to hit the mark, can at least show you that the situation is not entirely hopeless. It leaves unaltered Maxwell's equations, and indeed changes nothing in our views concerning the propagation of electromagnetic fields in the ether; but it introduces in the atoms themselves some compensating action such as could be performed by sufficiently intelligent demons of Maxwell.

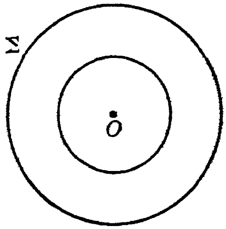

FIG. 28

Let O be the nucleus around which one or more electrons are revolving, and let Σ be some surface (Fig. 28) surrounding O at a short distance, — a sphere, for instance, so great that all the electrons of the atom are within it. The field at each point of Σ will change periodically in the course of time and can be decomposed by means of Fourier's theorem into a certain number of partial fields, of which the first is constant, whereas the others are represented by simple harmonic functions. Now for each of these partial fields we can imagine over the surface Σ a certain distribution of electromotive and perhaps magnetomotive forces, such that this system of forces, S, could produce the field considered at all outside points. Then it is clear that if we introduce at the surface Σ a system of forces equal and opposite to the system S, say the system $-S$, and if we do so for all the variable partial fields of which I have spoken, we shall destroy the radiation, keeping only the constant part of the field due to the revolving electrons, — for instance, the constant magnetic field which they

produce.* We must further remember that it is necessary not only to free the system from radiation but also to maintain its motion. For this purpose we imagine certain forces $-R$, acting on all the electrons and equal and opposite to the resistances R of which we have spoken.

In this way we have really found a suitable mechanism. There is no internal contradiction between our suppositions. We can imagine all the forces which we have now introduced to be exerted by some hidden natural system. This system would do a positive work by means of the forces $-R$ and an equal negative work by means of the forces $-S$, so that the total work of the hidden system is zero. Therefore it can continually perform the functions for which we have invented it.

29. Principles of the Theory of Special Relativity. Half a century ago physicists were much interested in the question whether the motion of translation which the earth has in its annual course around the sun can show itself in an influence on optical phenomena. The aberration of light, which was discovered by Bradley in the beginning of the eighteenth century, and which could be so easily explained by the emission theory, had led Fresnel to the conclusion that the ether must be considered as absolutely immovable, all ponderable bodies, even when of the size of planets, being perfectly permeable to it. This theory of a stationary ether had been applied with success to problems concerning a possible influence of the earth's motion on optical phenomena. Experience had never shown any change, and this negative result was easily explained by theory, so far at least as quantities of the second order, with respect to the ratio between v, the velocity of the earth, and c, the velocity of light, could be neglected.

A few remarks must be made on the methods by which the theoretical result was obtained. When a body like the earth moves with a constant velocity v in the direction, say, of OZ, it is natural to introduce a set of axes of coördinates that are moving with the body. If the coördinates with respect to them

* See Note 8, Appendix.

are represented by x', y', z', whereas x, y, z are the coördinates with respect to axes at rest in the ether, we shall have

$$x' = x, \quad y' = y, \quad z' = z - vt \tag{114}$$

as the formulæ of transformation.

In problems of ordinary mechanics the equations describing phenomena are much the same whether we use x, y, z or x', y', z'. Suppose, for instance, that a point (x, y, z) moves with uniformly accelerated motion,

$$\dot{x} = 0, \quad \ddot{y} = 0, \quad \ddot{z} = a,$$

then the point x', y', z' also moves with uniformly accelerated motion,

$$\ddot{x}' = 0, \quad \dot{y}' = 0, \quad \ddot{z}' = a.$$

When one tried to use a similar substitution for electromagnetic phenomena, it was found that if the equations were to maintain their form, it was necessary to introduce not only a new coördinate in the line of motion but also a new variable t' instead of t, so that to (114) one had to add

$$t' = t - \frac{v}{c^2} z,$$

where c is the velocity of light in a vacuum.

So far the theory indicates that no changes in optical phenomena depending on the first power of v/c should be observed, but that changes depending on squares and higher powers of v/c would not be excluded. Since v/c is about 0.0001, its square is a very small quantity; nevertheless, in one case there seemed to be a possibility of detecting a second-order effect. This was in Professor Michelson's classical interference experiment.

30. The Michelson-Morley Experiment. The underlying idea of this experiment is due to Clerk Maxwell. Suppose A and B to be two points of the earth such that the line AB is in the direction of the earth's motion of translation. We suppose the ether to be stationary; then by the translation of the earth we mean its translation with velocity v through or with respect to the ether.

A beam of light is now made to pass from A to B and back again. The velocity of the light relative to the earth is in one case $c+v$ and in the other $c-v$. If $AB=l$, the time required is

$$\frac{l}{c+v}+\frac{l}{c-v}=\frac{2\,cl}{c^2-v^2},$$

or, since v^2/c^2 is very small,

$$\frac{2\,l}{c}+2\,l\frac{v^2}{c^3}. \tag{115}$$

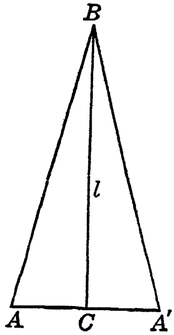

FIG. 29

There is a similar problem in which the earth's velocity v is at right angles to AB. If the diagram in Fig. 29 is at rest in the ether, ABA' will be the path of the beam, where $AB : AC = c : v$. B and C are simultaneous positions of B and A respectively; consequently $BC = l$.

Hence, $$AB=\frac{cl}{\sqrt{c^2-v^2}};$$

the time required for the double journey is thus

$$\frac{2\,l}{\sqrt{c^2-v^2}}=\frac{2\,l}{c}+l\frac{v^2}{c^3}. \tag{116}$$

In the actual experiment a beam of light was divided into two parts by a half-silvered plate of glass M. The two beams were reflected at the mirrors A and B (Fig. 30), and the second was reflected again at the plate M. The beams were now approximately parallel, and interference fringes were produced at C.

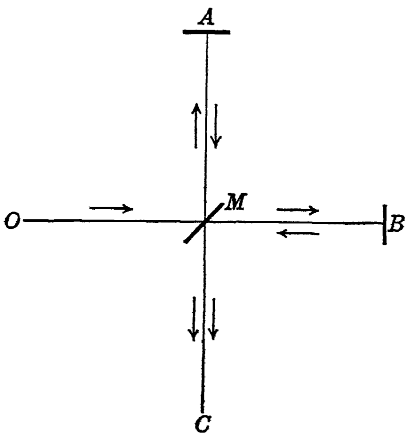

FIG. 30

The apparatus was first set up so that OB was in the direction of the earth's motion and then rotated slowly into a new position

in which MA was in the direction of the earth's motion. The difference between the times (115) and (116) would lead us to expect a displacement of the interference fringes, but the result of the experiment was negative. To explain the negative result the hypothesis that the apparatus contracts in the direction of the earth's translation in the ratio $c : \sqrt{c^2 - v^2}$ was put forward by Fitzgerald* and myself† in 1892.

Then arose the important question whether as a rule all optical and electromagnetic phenomena would be found to be independent of the earth's motion, even when squares and perhaps higher powers of v/c were taken into account, — in other words, for any value of v/c less than unity.

31. The Relativity Transformation. This led to the introduction of new independent variables x', y', z', t', defined by formulæ different from the above equations; namely,

$$x' = x, \quad y' = y, \quad z' = \frac{z - vt}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad t' = \frac{t - \frac{vz}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{117}$$

If the introduction of these variables, together with suitably chosen vectors $\mathbf{E}'$ and $\mathbf{H}'$ instead of the former vectors $\mathbf{E}$ and $\mathbf{H}$ of the electromagnetic field, leaves the field equations in their old form, the equations can be said to be invariant, or, as is often said, covariant. I did not succeed in proving the general covariancy of the field equations, but Einstein did.‡ Moreover, he inverted the method. Instead of starting from known equations and proving that for these covariancy exists, he states the postulate that covariancy shall exist and uses this for deductions concerning the form of the equations; that is, concerning the laws of force and properties of matter that are expressed by them.

* See Oliver Lodge, "Aberration Problems," *Phil. Trans. A*, Vol. 184 (London, 1893), p. 727.

† Abhandlungen über theoretische Physik, Bd. I, p. 441.

‡ *Ann. d. Phys.*, Bd. 17 (1905), p. 891.

We shall now follow this course. Indeed, if you want to get familiar with the theory of relativity, you can do no better than to exercise yourself in drawing deductions from the fundamental postulate.

We shall write the equations in the abbreviated form

$$x' = x, \quad y' = y, \quad z' = az - bct, \quad t' = at - \frac{b}{c}z \tag{118}$$

by putting
$$a = \frac{c}{\sqrt{c^2 - v^2}}, \quad b = \frac{v}{\sqrt{c^2 - v^2}}, \tag{119}$$

so that
$$a^2 - b^2 = 1. \tag{120}$$

We shall assume that a is positive; b will be positive or negative as the case may be.

We shall call this the relativity transformation. Of course there is a similar transformation for any arbitrarily chosen direction instead of Ox, but this one will suffice. Some simple deductions from these equations will now be given.

1. If we solve the equations for x, y, z, t, we get

$$x = x', \quad y = y', \quad z = az' + bct', \quad t = at' + \frac{b}{c}z'. \tag{121}$$

We can interchange the symbols with and without primes if at the same time we change the sign of b. This remark applies to all the formulæ of the theory of relativity.

2. Any phenomenon whatever can be described in terms of the coördinates x, y, z and the time t, or, as we shall also say, in terms of the four coördinates x, y, z, t of the first system, and just as well in terms of the coördinates x', y', z', t'. In the second system t' plays the part of the time. That one can use different times is natural; in ordinary life you can go by the Pacific time, or the Mountain time, or any local time you please, and we can imagine that the unit of time might be somewhat different in Chicago from what it is here in Pasadena. The numerical relation between the numerical values of two local times, which is now simply the constancy of their difference, would then be of a more complicated nature, something like our relation

$$t' = at - \frac{b}{c}z. \tag{122}$$

3. Let us consider a point of the second system for which we have constantly $x' = 0$, $y' = 0$, $z' = 0$. For this point

$$az - bct = 0, \quad \text{or} \quad z = \frac{b}{a} ct;$$

that is, the origin of coördinates in the second system has in the first system a constant velocity in the direction of OZ. If we call this velocity v, we have

$$\frac{b}{a} = \frac{v}{c},$$

which, combined with $a^2 - b^2 = 1$, gives

$$a = \frac{c}{\sqrt{c^2 - v^2}}, \quad b = \frac{v}{\sqrt{c^2 - v^2}}.$$

Similarly, the motion of the origin of coördinates of the first system, considered in the second system, is found by putting $z = 0$ in

$$z = az' + bct'.$$

This leads to $\quad 0 = az' + bct', \quad z' = -\frac{b}{a} ct',$

so that the motion is uniform but in the direction opposite to that of the motion just considered. The velocity is of the same magnitude.

32. Deductions from the Principle of Relativity. Suppose that we know a certain phenomenon *I* which we can describe in terms of x, y, z, t. Then, according to the principle of relativity, there must be another phenomenon which is described in exactly the same way in terms of x', y', z', t'. Using the transformation formulæ, we can describe this phenomenon also in terms of x, y, z, t; the equations to which we are led in this way will be different from those from which we started. So, by means of the principle, we can, knowing a certain phenomenon, conclude the possibility of another, *II*.

Example. In phenomenon *I* a rod having the direction of the axis of z has its extremities at the points

$$z_1 = p, \quad z_2 = p + l,$$

p and l being constant. Thus the rod is at rest, and its length, measured by the difference of the values of z, is l.

We conclude the possibility of the case

$$z_1' = p, \quad z_2' = p + l,$$

or according to the transformation formulæ (which we can apply separately to both ends),

$$az_1 - bct = p, \quad az_2 - bct = p + l,$$

$$z_1 = \frac{p}{a} + \frac{bc}{a}t, \quad z_2 = \frac{p}{a} + \frac{l}{a} + \frac{bc}{a}t.$$

This is case *II*. The rod is moving with velocity $v \left(= \frac{b}{a}c\right)$ in the direction of its length, and its length, measured in exactly the same way as that of the stationary rod, is

$$\frac{l}{a} = l\sqrt{1 - \frac{v^2}{c^2}}.$$

This is the contraction of which we have already spoken.

The way in which one phenomenon is deduced from another by means of the principle of relativity in this example may be characterized by the statement

$$II(\text{described in } x', y', z', t') = I(\text{described in } x, y, z, t).$$

But we can also reason in a somewhat different way (though it amounts to the same thing), beginning by describing the phenomenon *I*, which is given in terms of x, y, z, t, in terms of x', y', z', t'. Our new phenomenon will then be the same in x, y, z, t as *I* was in x', y', z', t'.

If, as we shall often do, we imagine two observers, A and B, the first of whom describes phenomena in terms of x, y, z, t, whereas the second uses x', y', z', t' (we can think of them, though this is not necessary, as occupying fixed places, the one in x, y, z, the other in x', y', z'), we can say that the new phenomenon will be one that presents itself to A just as *I* presents itself to B. Thus we have only to describe phenomenon *I* in x', y', z', t' and find out how it appears then (that is, to the observer B). We need not repeat every time that the new phenomenon is one that has the same aspect for A, and we

need not write down its equations, which are obtained by simply dropping all primes in the formulæ by which I is described in x', y', z', t'.

Some further examples may illustrate what has been said.

1. Suppose that we have a material system which is so small that we can characterize its position as a whole by definite values of x, y, z. Suppose, further, that in this system certain phenomena occur, succeeding each other at equal intervals of time. The system may, for instance, be a clock, and the phenomena in question the successive ticks.

If, now, by means of the formulæ of transformation, we calculate the corresponding values of x', y', z', t', we shall find the description of the same event in the new system of coördinates; namely, a specification of the place and time in x', y', z', t'.

Consider a clock at rest in x, y, z, t; let the interval between two consecutive ticks be s. Take this as our phenomenon I and describe it by means of the equations

$$x = \alpha, \quad y = \beta, \quad z = \gamma \ (\alpha, \beta, \gamma \text{ constants});$$
$$t = t_0, \quad t_0 + s, \quad t_0 + 2\,s, \text{ etc.}$$

Transforming to x', y', z', t', we get

$$x' = \alpha, \quad y' = \beta, \quad z' = \frac{\gamma}{a} - \frac{b}{a}ct',$$

$$t' = at_0 - \frac{b\gamma}{c}, \quad at_0 + as - \frac{b\gamma}{c}, \quad at_0 + 2\,as - \frac{b\gamma}{c}, \text{ etc.}$$

In this system the clock is moving with velocity $-\frac{b}{a}c = v$, say, and the intervals between the ticks are no longer s but are somewhat greater than s. Thus, if the clock has a motion of translation with velocity v, it goes slower in a ratio determined by $\sqrt{1 - \frac{v^2}{c^2}}$.

2. Suppose that a light-signal is made to pass from one end P of a rod to the other end Q and then back to P. Let the interval between the instants of starting and returning be τ, the rod being supposed to be at rest. Then the same phenomenon described in x', y', z', t' will be as follows:

The rod has a motion of translation, and the interval of time between the start and return of the light-signal is (in virtue of the equations) equal to $a\tau$. Thus we have a new phenomenon; in fact, we find that when the rod moves with velocity v, the time taken for

light to travel along it to and fro will be greater than when the rod is at rest in the ratio of $\sqrt{1-\frac{v^2}{c^2}}$ to 1. The interesting thing about this is that this change is independent of the direction of the rod, and here we have the explanation of Michelson's experiment.

That the negative result of the experiment is implied in the principle can also be seen as follows: Suppose that in Michelson's experiment two fine micrometer wires are set on the places where we have darkness; that is, at consecutive minima. Then, if we describe this in x', y', z', t', we shall have a system in motion but still no light at the places occupied by the wires.

3. We may remark here that when we say that in the systems x, y, z, t and x', y', z', t' natural phenomena are represented by equations of the same form, this means in the case of light that its propagation can be described as taking place with velocity c, the same in both cases. The proof is as follows:

In the system x, y, z, t a light-signal goes from the point 0, 0, 0 at time $t=0$ to a point x, y, z which it reaches at time t. Describing this in terms of x', y', z', t', we can say that the signal goes from 0, 0, 0 at time 0 to a point x', y', z' which it reaches at time t', where

$$x'=x, \quad y'=y, \quad z'=az-bct, \quad t'=at-\frac{b}{c}z.$$

From these we deduce

$$z'+ct'=(a-b)(z+ct), \quad z'-ct'=(a+b)(z-ct).$$

Multiplying, we get $\qquad z'^2-c^2t'^2=z^2-c^2t^2,$

and therefore $x'^2+y'^2+z'^2-c^2t'^2=x^2+y^2+z^2-c^2t^2.$

Consequently, since $\qquad t=\frac{1}{c}\sqrt{x^2+y^2+z^2},$

we have also $\qquad t'=\frac{1}{c}\sqrt{x'^2+y'^2+z'^2}.$

This means that in the system x', y', z', t' we have also to ascribe to light the velocity c.

Having gone thus far, we may again conclude that the result of Michelson's experiment must be negative.

Let me now give you the following picture:

Suppose that we have two systems of rods all having the direction of the axis of z and all equal to each other, one system being at rest in the system x, y, z, t, the other in the

system x', y', z', t'. We may call the first the A-rods and the second the B-rods, and we may consider them to belong to two observers A and B who are themselves at rest in the two systems.

Suppose that we have also two systems of clocks, A-clocks and B-clocks, at rest respectively in the two systems. The A-rods, if properly juxtaposed, will give us for each point of space a value of z, and the B-rods a value of z'. We can have numbers written at the points where the rods touch each other, these numbers indicating the values of z or of z', as the case may be. We can also have rods in the directions of OX, OY (OX', OY'), so that we have two scaffoldings of rods, on which we can read off immediately values of x, y, z or x', y', z'. These two systems of rods are interpenetrating, and we shall imagine them to be able, notwithstanding this, to move relatively to each other with velocities $v = -\frac{b}{a}c$ and $v = \frac{b}{a}c$, which we know already. We can further imagine that the A-clocks and the B-clocks are fixed to these structures, and that there is still sufficient room for all kinds of phenomena in spite of the existence of all these rods and clocks. We shall suppose that all the A-clocks are adjusted to each other with the utmost precision. In order to do so we send a light-signal from one A-clock to another, and we give the hands such positions that the distance of these positions, namely, of the first clock at the moment when the signal is sent and of the second clock at the moment when the signal arrives, is exactly equal to r/c, the distance r being determined by means of the A-rods. In exactly the same way all the B-clocks may be adjusted to each other.

If the number of rods and clocks is sufficiently large, we can read off for any event the values of x, y, z, t and those of x', y', z', t', and between them one will find everywhere the relations that are expressed by the transformation formulæ, and by looking attentively at the system one would find comprised therein what I have said about the contraction of moving rods and the retardation of moving clocks.

It is important to remark that, if we think once more of our two observers A and B, the phenomena observed by B must obey exactly the same laws as those witnessed by A. Thus, for B also, a moving rod must be shortened and a moving clock must go more slowly than a stationary one. Now whereas, from the point of view of the observer A, the A-rods are stationary and the B-rods are moving, for the observer B the reverse will be the case, and so for him the A-rods will be shorter than the B-rods. Similarly, he will conclude that the A-clocks are going more slowly than the B-clocks.

So we come to the conclusion that of two rods the first may be held to be the shorter by one observer and the second by another observer. There seems something paradoxical about this, but in reality there is nothing strange. For indeed, when A and B, each from his own point of view, compare the lengths of the two rods, they do not compare them in the same way. Of course, when we speak of making the comparison, we mean that we compare the lengths of the rods, measured by the difference between the coördinates of their extremities, such as these coördinates are at one and the same time. But for A the time is the variable t, while for B it is t'. Therefore A fixes his attention on the positions of the rods for a definite value of t, and B does so for a definite value of t'. This is not the same thing, and this explains the different results that are found in the two cases and that can be deduced from our formulæ.

I shall dwell no longer upon this point, but shall merely remark that, as indeed you can see at once from the equation

$$t' = at - \frac{b}{c} z,$$

if you consider events occurring at different places specified by different values of z, the values of t' are different when those of t are equal, and conversely. In other words, what is simultaneous in one system of coördinates is not simultaneous in the other system.

It may even happen that of two phenomena 1 and 2 the first is prior to the second in the system x, y, z, t and posterior to it

in x', y', z', t'. In fact, if we denote by suffixes 1 and 2 quantities relating to the two events, we have, by virtue of (118),

$$t'_2 - t'_1 = a(t_2 - t_1) - \frac{b}{c}(z_2 - z_1).$$

Let us suppose that b and $t_2 - t_1$ are positive; then $t'_2 - t'_1$ can be negative if $z_2 - z_1$ has a sufficiently great positive value. The condition is evidently

$$z_2 - z_1 > \frac{ac}{b}(t_2 - t_1),$$

and, since $a > b$, this implies that

$$z_2 - z_1 > c(t_2 - t_1);$$

that is, in the system in which 1 precedes 2 the distance between the points at which these events take place must be greater than the distance over which light travels in time $t_2 - t_1$.

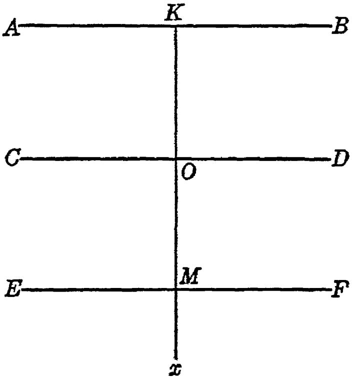

FIG. 31

I should like to emphasize the fact that the variations of length caused by a translation are real phenomena, no less than, for instance, the variations that are produced by changes of temperature. If our means of observation were refined enough, we should be able to observe the changes by means of instantaneous, say pinhole, photography, and if the experiment is first made by our observer A under proper circumstances, and then by B under circumstances that seem to him to be the same, it may very well be that of two rods the one has the shorter image for A and the other for B.

Let there be two "equal" rods P and Q, placed along the line AB parallel to OZ (our diagram, Fig. 31, is drawn in the system x, y, z), and let the first be at rest in x, y, z, t, the other in x', y', z', t'. The screen CD coincides with the plane YOZ, the distance OK being p, and EF is the photographic plate, at a distance $OM = q$ behind the screen.

In A's experiment the screen and the photographic plate are at rest in the system x, y, z, the pinhole being constantly at the origin O of those coördinates. We shall suppose it to be opened for an instant at time $t = 0$, and we shall define the "effective" position of a moving point by the condition that light emitted by the point in that position shall reach the hole O just when it is opened; the time at which the effective position is reached may be called the effective time. We shall denote by $2\,l$ the length of the rod P in the system x, y, z, so that in the same system $\frac{2}{a}l$ is the length of Q. For the sake of simplicity we shall suppose l to be so small that terms with l^2 may be neglected.

Now A must be sure that he compares *simultaneous* positions of the ends of the moving rod Q. He can satisfy this condition by arranging the experiment in such a way that the middle point of Q passes through K $(z = 0)$ at the time $-\frac{p}{c}$. We shall further assume that the middle point of P always has this position.

It is easily seen that in this first experiment the images of the two rods have the lengths $2\frac{q}{p}l$ and $2\frac{q}{p}\frac{l}{a}$, the second being a times shorter than the first.

Passing on to the second experiment, we remark, in the first place, that if it is to be for B exactly what the first experiment was for A, the screen and the photographic plate must be at rest in the system x', y', z', and the opening has to be uncovered at the time $t' = 0$, which is equivalent to $t = 0$ because, for the opening, z' must be 0. Furthermore, the middle point of Q must be continually at the point $z' = 0$, and the middle point of P must have that position at the instant $t' = -\frac{p}{c}$. We have, therefore, by the equations of transformation, for any value of t,

$$z = -bp$$

for the middle point of P and

$$z = \frac{b}{a}ct$$

for that of Q.

Starting from these data, we shall now make the calculation

for the second experiment, availing ourselves of the system x, y, z, t. In the first place, knowing that the lengths of the rods are $2\,l$ and $\frac{2}{a}l$, we find, for the coördinates of the extremities of P,

$$-l-bp, \quad l-bp,$$

and for those of the ends of Q,

$$-\frac{1}{a}l+\frac{b}{a}ct, \quad \frac{1}{a}l+\frac{b}{a}ct.$$

This leads * to the following values of the effective times

$$t_{1P}=-\frac{a}{c}p-\frac{b}{ac}l, \quad t_{2P}=-\frac{a}{c}p+\frac{b}{ac}l,$$

$$t_{1Q}=-\frac{a}{c}p-\frac{b}{c}l, \quad t_{2Q}=-\frac{a}{c}p+\frac{b}{c}l,$$

the effective positions being

$$z_{1P}=-l-bp, \quad z_{2P}=l-bp,$$

$$z_{1Q}=-al-bp, \quad z_{2Q}=al-bp.$$

Let $\bar{t}_{1P}$, $\bar{z}_{1P}$ etc. be the values of t and z characterizing the "events" of the formation of the four images. As the light travels along straight lines, passing through the point $z=0$ at the instant $t=0$, we shall find $\bar{t}_{1P}$, $\bar{z}_{1P}$, etc. if we multiply the above values by $-\frac{q}{p}$.

We have finally to take into account the fact that the photographic plate moves with the velocity $\frac{b}{a}c$, and that therefore, if impressions are made on it at $t=\bar{t}_1$, $z=\bar{z}_1$ and $t=\bar{t}_2$, $z=\bar{z}_2$, the values of z for simultaneous positions of the dots produced will differ by

$$\bar{z}_2-\bar{z}_1-\frac{b}{a}c(\bar{t}_2-\bar{t}_1).$$

In this way we find, for the lengths of the images of P and Q, measured by the difference of the values of z,

$$\frac{q}{p}\cdot\frac{2}{a^2}l \quad \text{and} \quad \frac{q}{p}\cdot\frac{2}{a}l,$$

* If the motion of a point is given by $z=\alpha+\beta t$, its distance from O is $\sqrt{p^2+(\alpha+\beta t)^2}$ and the condition for the effective position is that this shall be equal to $c\,|t|$. In the calculations l^2 is constantly neglected, and frequent use is made of the relation $a^2-b^2=1$.

so that now the second image is a times *longer* than the first.* The influence of a translation on the rate of motion of a clock is no less real than the shortening of a rod.

Suppose we have two "equal" clocks P and Q, one of them having a fixed position in our system x, y, z, t, while the other is moving along a straight line from A to some distant point B and back to A. We shall suppose that the velocity increases from zero to a value v during an interval of time τ_1, that it remains constant during τ_2, is inverted during τ_3, and remains constant during τ_4, and that finally the clock is brought to rest in the interval τ_5. Then, surely, we may expect that, on comparing the two clocks, we shall see the effect of this wandering.

Of course, since we do not know what will happen to the clock in the intervals of time τ_1, τ_3, τ_5, there would be an element of uncertainty in the experiment. This could, however, be eliminated by repeating it with only a change in the distance AB, so that the intervals τ_2 and τ_4 become longer or shorter.†

33. Transformation Formulæ for Velocities. Relation between Elements of Volume in the Two Systems. Suppose that the coördinates x, y, z of a moving point are given as functions of t. We can then express x', y', z' in terms of t' by means of the transformation formulæ; that is, from a description of the motion in x, y, z, t we can deduce that in x', y', z', t'.

What is the relation between the velocities? If dx, dy, dz are the changes in the coördinates x, y, z of the moving point in time dt, we have

$$v_x = \frac{dx}{dt}, \quad v_y = \frac{dy}{dt}, \quad v_z = \frac{dz}{dt}.$$

But to dx, dy, dz, dt correspond the changes

$$dx' = dx, \quad dy' = dy, \quad dz' = a\,dz - bc\,dt, \quad dt' = a\,dt - \frac{b}{c}\,dz,$$

* If we wanted to know the lengths of the two images measured on a scale that is drawn on the photographic plate, we should have to take into account the contraction of the scale caused by its translation. But it was not necessary to speak of this, because the ratio between the two lengths was all we required.

† See Note 9, Appendix.

and in the system x', y', z', t' we have to consider as the components of velocity the quantities

$$v'_x = \frac{dx'}{dt'} = \frac{dx}{a\,dt - \frac{b}{c}\,dz} = \frac{v_x}{a - \frac{b}{c}v_z},$$

$$v'_y = \frac{dy'}{dt'} = \frac{dy}{a\,dt - \frac{b}{c}\,dz} = \frac{v_y}{a - \frac{b}{c}v_z},$$

$$v'_z = \frac{dz'}{dt'} = \frac{a\,dz - bc\,dt}{a\,dt - \frac{b}{c}\,dz} = \frac{av_z - bc}{a - \frac{b}{c}v_z},$$

or

$$v'_x = \frac{v_x}{\omega}, \quad v'_y = \frac{v_y}{\omega}, \quad v'_z = \frac{av_z - bc}{\omega}, \tag{123}$$

where

$$a - \frac{b}{c}v_z = \omega. \tag{124}$$

We can also express v_x, v_y, v_z in terms of v'_x, v'_y, v'_z by means of formulæ similar to the above except that the sign of b is changed. In these formulæ the denominator will be the quantity

$$a + \frac{b}{c}v'_z = \omega'. \tag{125}$$

It is interesting to note the following consequences of these relations:

$$c^2 - v'^2 = \frac{c^2 - v^2}{\omega^2}, \quad c^2 - v^2 = \frac{c^2 - v'^2}{\omega'^2}, \quad \omega' = \frac{1}{\omega}. \tag{126}$$

We may further remark that there are three velocities, namely v, v', and the velocity $\frac{b}{a}c$, with which one system of coördinates moves with respect to the other. Since $b < a$, the last of these velocities is always smaller than c. We shall also suppose that v and v' are always smaller than c. This leads us to no contradiction, since, according to the above formulæ, $v < c$ implies $v' < c$, and conversely. Indeed, one of the fundamental theorems of the theory of relativity is that a body can never have a velocity greater than the speed of light c. The reason for the exclusion of velocities higher than c will be stated later on, but

for the present an example may serve to make clear how the theorem is to be understood. Suppose that in the system x, y, z, t we have two particles moving in the direction OZ with velocities αc and $-\alpha c$, α being a fraction that can be near 1. Then the difference of the velocities $2\alpha c$ can very well be greater than c, and if you want to call this difference the relative velocity of one particle with respect to the other (to which there is no objection), you can say that this velocity is greater than c. What is meant in the statement just made is that the velocity of a particle in a system of coördinates, introduced by means of a relativity transformation, can never be greater than c if it was less than c in the original system.

If, for instance, we take a set of axes moving with one of the particles, the constants a and b of the transformation to the new set of coördinates will be

$$\frac{1}{\sqrt{1-\alpha^2}} \quad \text{and} \quad -\frac{\alpha}{\sqrt{1-\alpha^2}},$$

and the velocities of the two particles in the new system will be $\dfrac{2\,\alpha c}{1+\alpha^2}$ and 0 respectively. It should be noticed that the first quantity is always less than c.

In general the velocity is changed by the transformation; that is,

$$|v| \neq |v'|;$$

but if $v^2 = c^2$, $v'^2 = c^2$ likewise.

That the velocity of light c presents the special feature that it is not changed by a relativity transformation is, of course, due to the fact that c enters in the transformation formulæ. If, moreover, the transformation which they determine is such as to leave the laws of physical phenomena unchanged, this must mean that in all these phenomena something with which we are concerned in the propagation of light plays a part. So the principle of relativity implies a certain relationship between optical or electromagnetic effects and other physical phenomena.

Before proceeding farther I must make you acquainted with a formula that is often used in relativity. Suppose that we have a system of points like the molecules of a gas but moving

in such a way that the velocities do not change abruptly and irregularly, as they would in a gas, but so that they are continuous functions of the coördinates and of the time. Let P be one of these points having the coördinates $\bar{x}, \bar{y}, \bar{z}$ at time $\bar{t}$, and let dV be an element of volume in the space x, y, z surrounding P. The magnitude of dV is evaluated on the basis that the magnitude of an element having the form of an infinitely small parallelepiped with sides dx, dy, dz is taken to be $dx\,dy\,dz$. In the element dV we shall find at time $\bar{t}$ a definite part of our system of points.

For each point we can now seek the values of x', y', z' such as they are for a definite value, say $\bar{t}'$, of the new time, and all the positions that we find in this way will be included in a definite element dV' in the space x', y', z'.

For $\bar{t}'$ we shall take the value of t' which according to the transformation formulæ corresponds to $\bar{x}, \bar{y}, \bar{z}, \bar{t}$, that is, to the coördinates of P and the moment $\bar{t}$ which we want to consider; and we shall now compare the magnitude of dV' with that of dV, determining the first of these by the rule that in the space x', y', z' an infinitely small parallelepiped with sides dx', dy', dz' has the magnitude $dx'dy'dz'$.

You see we want to compare the spaces over which, in the two systems, simultaneous positions of the particles are spaced, *simultaneous* meaning that in one case t, and in the other t', has a definite value.

Let $\bar{x}+x, \bar{y}+y, \bar{z}+z$ be the coördinates at time t of some particle Q lying within dV, x, y, and z being infinitely small. The corresponding coördinates in the system x', y', z', t' are

$$\bar{x}+x, \quad \bar{y}+y, \quad a(\bar{z}+z)-bc\bar{t}, \quad a\bar{t}-\frac{b}{c}(\bar{z}+z);$$

that is, the first three give us the x', y', z' coördinates of Q at time $a\bar{t}-\frac{b}{c}(\bar{z}+z)$; but we want to know the values of these coördinates at another time $\bar{t}'=a\bar{t}-\frac{b}{c}\bar{z}$, the difference in time being $\frac{b}{c}z$. In this time the coördinates have changed by

$$\frac{b}{c}zv'_x, \quad \frac{b}{c}zv'_y, \quad \frac{b}{c}zv'_z,$$

where v'_x, v'_y, v'_z are the components of the velocity v' of the particle Q. But, neglecting quantities of the second order, we may just as well understand by these symbols the velocity of the particle P. Thus the desired coördinates of Q are

$$x' = \bar{x} + x + \frac{b}{c} z v'_x, \quad y' = \bar{y} + y + \frac{b}{c} z v'_y,$$
$$z' = a(\bar{z} + z) - bc\bar{t} + \frac{b}{c} z v'_z.$$

When x, y, z take all values consistent with positions of Q lying within dV, the point x', y', z' occupies successively all possible positions in dV', and according to a well-known geometrical theorem the ratio of dV and dV' is given by what is called the functional determinant of x', y', z' with respect to x, y, z; that is, the determinant whose elements are the partial derivatives of x', y', z' with respect to x, y, z. Thus

$$\frac{dV'}{dV} = \begin{vmatrix} 1 & 0 & \frac{b}{c} v'_x \\ 0 & 1 & \frac{b}{c} v'_y \\ 0 & 0 & a + \frac{b}{c} v'_z \end{vmatrix},$$

or, on account of (125) and (126),

$$\frac{dV}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{dV'}{\sqrt{1 - \frac{v'^2}{c^2}}}. \tag{127}$$

34. Relativity and the Electromagnetic Equations. The principle of relativity is a physical principle, or a physical hypothesis, which pretends to teach us something about the nature of things. The consequences to which it leads have to be tested experimentally, and when these consequences concern phenomena that we are accustomed to explain by some theory, the principle of relativity may imply some corresponding change in that theory.

For instance, it is natural to consider the length of the rod and the elasticity of the spring that regulates the motion of a

clock as determined by molecular forces. The principle of relativity requires that these forces should be modified by motion. If we had known this modification beforehand, we should have been able to deduce from it the contraction of rods and the retardation of clocks; it would, in fact, have been possible to deduce the principle of relativity for the class of phenomena in question, and this is the course that has really been followed in the case of the electromagnetic field.

The equations of Maxwell do not change their form when the relativity transformation is applied. In fact, if we put

$$\left.\begin{array}{lll} E'_x = aE_x - bH_y, & E'_y = aE_y + bH_x, & E'_z = E_z \\ H'_x = aH_x + bE_y, & H'_y = aH_y - bE_x, & H'_z = H_z \end{array}\right\}, \tag{128}$$

$$\rho' = \omega\rho, \tag{129}$$

the equations in the new variables have the same form as the original ones. This is easily verified if it is taken into account that, in virtue of the equations of transformation (121),

$$\frac{\partial}{\partial x'} = \frac{\partial}{\partial x}, \quad \frac{\partial}{\partial y'} = \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial z'} = a\frac{\partial}{\partial z} + \frac{b}{c}\frac{\partial}{\partial t}, \quad \frac{\partial}{\partial t'} = a\frac{\partial}{\partial t} + bc\frac{\partial}{\partial z}.$$

Thus, for instance,

$$\frac{\partial E'_x}{\partial x'} + \frac{\partial E'_y}{\partial y'} + \frac{\partial E'_z}{\partial z'} = \frac{\partial}{\partial x}(aE_x - bH_y) + \frac{\partial}{\partial y}(aE_y + bH_x)$$
$$+ \left(a\frac{\partial}{\partial z} + \frac{b}{c}\frac{\partial}{\partial t}\right)E_z = a \operatorname{div} \mathbf{E} + b\left(\frac{\partial H_x}{\partial y} - \frac{\partial H_y}{\partial x}\right) + \frac{b}{c}\frac{\partial E_z}{\partial t}.$$

Substituting $\operatorname{div} \mathbf{E} = \rho$ and $\dfrac{\partial H_y}{\partial x} - \dfrac{\partial H_x}{\partial y} = \dfrac{1}{c}\left(\dfrac{\partial E_z}{\partial t} + \rho v_z\right)$,

this becomes $$a\rho - \frac{b}{c}\rho v_z = \omega\rho = \rho'.$$

Similarly for the other equations.

Equation (129) is found when one considers the electric charges as concentrated in a great number of points (which may be exceedingly close to each other), the charge of each point being invariable and being expressed in the same unit, whether we use the system x, y, z, t or x', y', z', t'. On these assumptions

the elements dV and dV', of which we spoke in § 33, must contain the same total charge. Thus

$$\rho' dV' = \rho \, dV, \quad \rho' = \rho \frac{dV}{dV'} = \omega\rho.$$

This equation shows (and the same is seen for instance in formula (122) and those which we found for the velocities) that the two observers A and B, to revert to them once more, do not introduce into their theories quantities that are equal, but rather corresponding quantities, mutually connected by definite transformation formulæ and such that in the two modes of description they have similar meanings. Thus, while **E** is the force acting on a unit charge of electricity that is at rest in the system x, y, z, t, the force acting on electricity that is at rest in x', y', z', t' is obtained by multiplying the charge by **E**′. Since in this case the electricity is in motion in the x, y, z, t system, and therefore experiences a force due to the magnetic field, it is not strange that **E**′ is connected not only with **E** but also with **H**.

35. Zeeman's Experiments on the Propagation of Light in Moving Solid Bodies. Consider a rod of glass R (Fig. 32) at rest in the x, y, z, t system and extending from $z = 0$ to $z = l$.

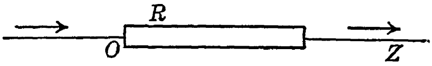

FIG. 32

A beam of light, in which the vibrations are represented by expressions containing the factor

$$\cos n\left(t - \frac{z}{c}\right) \tag{130}$$

is incident on the left-hand side. In the rod the corresponding factor is

$$\cos n\left(t - \frac{z}{v_n}\right) = \cos n\left(t - \frac{\mu_n z}{c}\right), \tag{131}$$

where v_n is the velocity of light for frequency n and μ_n is the index of refraction for this frequency. No constant phase term is added to $t - \frac{\mu_n z}{c}$, because at $z = 0$ the phase of the incident waves and of the waves in the glass must be the same.

On the right-hand side of the rod the vibrations in the ether are represented by terms containing the factor

$$\cos n\left[t - \frac{z}{c} - (\mu_n - 1)\frac{l}{c}\right], \tag{132}$$

the constant term within the brackets being chosen so that for $z = l$ the phase agrees with that in the rod.

Now by means of the principle of relativity we may deduce the laws of a new phenomenon; namely, the propagation of light in a moving rod of glass. As we shall be concerned only with the incident and emergent light, we shall consider only (130) and (132), not (131).

In the second phenomenon we have, in the system x', y', z', t', incident rays represented by

$$\cos n\left(t' - \frac{z'}{c}\right) \tag{133}$$

and emergent rays represented by

$$\cos n\left[t' - \frac{z'}{c} - (\mu_n - 1)\frac{l}{c}\right]. \tag{134}$$

We now express these in terms of x, y, z, t. In this system of coördinates the rod has the velocity $w = \frac{b}{a}c$.

If we neglect quantities of order $\frac{w^2}{c^2}$, we can write $a = 1$, $b = \frac{w}{c}$, so that the formulæ of transformation become

$$z' = z - wt, \quad t' = t - \frac{w}{c^2}z.$$

Thus the incident light is represented by

$$\cos n\left[t - \frac{w}{c^2}z - \frac{z - wt}{c}\right],$$

and the emergent light by

$$\cos n\left[t - \frac{w}{c^2}z - \frac{z - wt}{c} - (\mu_n - 1)\frac{l}{c}\right],$$

or, again neglecting squares of $\frac{w}{c}$, the incident light is represented by

$$\cos \bar{n}\left(t - \frac{z}{c}\right), \tag{135}$$

and the emergent light by

$$\cos \bar{n}\left[t - \frac{z}{c} - (\mu_n - 1)(1 - \frac{w}{c})\,\frac{l}{c}\right], \tag{136}$$

where

$$\bar{n} = n\left(1 + \frac{w}{c}\right). \tag{137}$$

Now compare (130) and (132) with (135) and (136), after making such changes in (135) and (136) that in both cases the incident light has the same frequency. Because the difference between n and $\bar{n}$ is an infinitely small quantity of order w/c we can put in (136)

$$\mu_n = \mu_{\bar{n}} + \frac{d\mu}{dn}(n - \bar{n}) = \mu_{\bar{n}} - n\frac{d\mu}{dn}\cdot\frac{w}{c}, \tag{138}$$

so that (136) becomes

$$\cos \bar{n}\left[t - \frac{z}{c} - (\mu_{\bar{n}} - 1)\,\frac{l}{c} + \left(\mu_{\bar{n}} + n\frac{d\mu}{dn} - 1\right)\frac{wl}{c^2}\right]. \tag{139}$$

Now $\bar{n}$ can have any value, and we can therefore replace it by n in both (135) and (139). The result is that the cosine factor for the incident light is

$$\cos n\left(t - \frac{z}{c}\right),$$

and for the light that has passed through the rod it is

$$\cos n\left[t - \frac{z}{c} - (\mu - 1)\,\frac{l}{c} + \left(\mu + n\frac{d\mu}{dn} - 1\right)\frac{wl}{c^2}\right],$$

μ being written for $\mu_{\bar{n}}$.

We have now a case in which the moving rod receives a beam of light of exactly the same nature as that which first fell on the stationary rod. You see, then, that the influence of the motion manifests itself as an acceleration of the phase of the emergent light amounting to

$$\left(\mu + n\frac{d\mu}{dn} - 1\right)\frac{wl}{c^2}. \tag{140}$$

This has been verified by Zeeman,* and he was even able to show by his measurements that the influence of the term $n\frac{d\mu}{dn}$ is perceptible. I have treated the problem at some length in order to show you how we can deduce results without entering into the details of phenomena. We did not consider the wave motion in the rod at all. The principle of relativity is in this respect like the second law of thermodynamics.

I need hardly add that the theory of the well-known experiments of Fizeau and other physicists on the propagation of light in moving fluids, by which "Fresnel's coefficient" was verified, can be developed in much the same way.

36. Modification of the Laws of Dynamics. Let us now revert to the transformation formulæ for velocities,

$$v'_x = \frac{v_x}{\omega}, \quad v'_y = \frac{v_y}{\omega}, \quad v'_z = \frac{av_z - bc}{\omega},$$

and to the equation

$$\sqrt{\frac{c^2 - v^2}{c^2 - v'^2}} = \omega = a - \frac{b}{c} v_z$$

that was found in connection with them. We can write the equations in such a form that in each of them we have on one side quantities referred to x', y', z', t' and on the other quantities referred to x, y, z, t; namely,

$$\left.\begin{array}{ll} k'v'_x = kv_x, & k'v'_y = kv_y \\ k'v'_z = akv_z - bck, & k' = ak - \frac{b}{c} kv_z \end{array}\right\}, \tag{141}$$

where

$$k = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad k' = \frac{1}{\sqrt{1 - \frac{v'^2}{c^2}}}. \tag{142}$$

These equations will enable us to draw a very important conclusion. We have already remarked that the principle of relativity can only be true if the laws that really govern physical

* P. Zeeman, *Proc. Acad. Amsterdam*, Vol. 22 (1920), p. 462; P. Zeeman and Miss A. Snethlage, ibid., p. 512.

phenomena are in accordance with it. In other words, the principle can lead us to the discovery of something about these laws, and it may very well be that the principle obliges us to make slight changes in the form of physical laws.

This is found to be the case with the fundamental laws of dynamics, such as, for instance, the law that, notwithstanding the mutual action between its parts, the momentum and energy of a mechanical system remain constant. It should be remarked that momentum and energy, rather than mass, are to be considered as fundamental ideas.

The impact of two bodies — perfectly elastic spheres, for instance, moving with their centers along a straight line — is a simple example of the mutual action between two parts of a system. In former days the laws of such collisions were investigated experimentally, and it was gradually recognized that if to each of the bodies we assign a certain constant m of properly chosen magnitude, the algebraic sum of the two quantities mv remains constant in a collision, and likewise the sum of the quantities $\frac{1}{2}mv^2$. In other words, $m_1v_1 + m_2v_2$ and $\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2$ are unaltered by a collision, v_1 and v_2 being the velocities of the two bodies. So physicists were led to define the momentum mv and the kinetic energy $\frac{1}{2}mv^2$ of a body moving with velocity v. The coefficient m was called the mass of the body.

Of course the experiments on impact, by which these laws were supposed to have been found, could not have been made with great precision, and experimenters labored under the difficulty that there are no perfectly elastic bodies. Until about fifteen years ago, however, physicists certainly thought that the closer the bodies approached to perfectly elastic bodies, and the more accurately the experiments were made, the nearer we should come to a verification of the above laws.

The principle of relativity asserts that this is not so, and that we have to change these laws. It is true that we can maintain the principles of the conservation of momentum and energy, but only on the condition that we change to a certain extent the definitions of momentum and energy. Indeed, the theory of relativity requires that the principles of the conservation of

momentum and energy must hold in the system x', y', z', t' as well as in x, y, z, t and of course also that momentum and energy must depend on the velocity in the same way in both systems of coördinates. Now if, considering the motion of two balls, say along the axis Oz, you could determine the velocities $\bar{v}_1, \bar{v}_2$ after collision by means of the formulæ

$$m_1\bar{v}_1 + m_2\bar{v}_2 = m_1 v_1 + m_2 v_2,$$
$$\tfrac{1}{2}\, m_1\bar{v}_1^2 + \tfrac{1}{2}\, m_2\bar{v}_2^2 = \tfrac{1}{2}\, m_1 v_1^2 + \tfrac{1}{2}\, m_2 v_2^2,$$

you could deduce from these, by means of the transformation formulæ, relations between the initial and final velocities expressed in terms of x', y', z', t'. But on account of the rather complicated transformation formulæ these new relations would be far from having the above form. It thus appears that the old definitions of energy and momentum cannot be maintained. How must we change them? The answer is given by equations (141). We define the energy by the equation

$$\epsilon = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} = kmc^2 \tag{143}$$

and the components of momentum by

$$g_x = kmv_x, \quad g_y = kmv_y, \quad g_z = kmv_z, \tag{144}$$

m being a constant factor for each body. Let us give to this factor the same value, whether we use the system x, y, z, t or x', y', z', t'. Then, according to (141),

$$g'_x = g_x, \quad g'_y = g_y, \quad g'_z = ag_z - \frac{b}{c}\,\epsilon, \quad \epsilon' = a\epsilon - bcg_z.$$

Now consider two colliding spheres. For each of them we have relations of the above form between momenta and energy in the systems x, y, z, t and x', y', z', t'. Therefore the same relations hold when by g and ϵ we understand the sums of momenta and energies of the bodies taken together. The same relations will, furthermore, hold also for changes in these total quantities; accordingly

$$\Delta g'_x = \Delta g_x, \quad \Delta g'_y = \Delta g_y, \quad \Delta g'_z = a\Delta g_z - \frac{b}{c}\Delta\epsilon, \quad \Delta\epsilon' = a\Delta\epsilon - bc\Delta g_z.$$

Hence if, during a collision (which now need not be a central one) or other mutual action, g_x, g_y, g_z, and ϵ do not change, g'_x, g'_y, g'_z, and ϵ' will likewise remain constant; and since these expressions depend in the same way on the respective components of velocity, the principle of relativity is satisfied.

These are the reasons for defining momentum and energy by the above formulæ. We can very well imagine that by very accurate experiments on collisions the laws of conservation of *these* momenta and energies are verified. These experiments would also give us the ratios between the constant coefficients m which we have to ascribe to different bodies, and which we consider as the measure of their masses. For these ratios we should find exactly the same values, whether we use x, y, z, t or x', y', z', t'.

You see that if the velocity of a body is small in comparison with that of light, the values now adopted for g_x, g_y, g_z differ only by small amounts from the classical values. As to the energy, things are different. We wrote

$$\epsilon = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Expanding the denominator by the binomial theorem, we find

$$\epsilon = mc^2 + \tfrac{1}{2}\, mv^2 + \cdots,$$

so that, apart from a constant term, we are led back to the old expression for kinetic energy.

In mechanics and thermodynamics we are always concerned with changes of energy, so that the constant term does not matter. We could drop it or replace it by an arbitrary constant, but then the transformation formulæ for g and ϵ would have to be modified.

Einstein attributes the above value to the energy, and this has certain advantages which will be noticed later on.

Having defined the momentum and energy of a particle, we can now consider the case when a force acts on the particle.

The force is measured by the corresponding rate of change of momentum; that is, its components are

$$F_x = \frac{d}{dt}\frac{mv_x}{\sqrt{1-\frac{v^2}{c^2}}}, \quad F_y = \frac{d}{dt}\frac{mv_y}{\sqrt{1-\frac{v^2}{c^2}}}, \quad F_z = \frac{d}{dt}\frac{mv_z}{\sqrt{1-\frac{v^2}{c^2}}}. \quad (145)$$

From these equations we can deduce the theorem that the work done by the force is equal to the change of energy. The rate at which the force does work is, in fact,

$$F_x v_x + F_y v_y + F_z v_z,$$

and this is found to be equal to

$$\frac{d}{dt}\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}. \quad (146)$$

37. Energy and Stresses etc. in Continuous Systems. From the equations concerning the momentum and energy of a particle we can pass on to the corresponding equations for systems that are distributed over space. What I am going to say now applies, for instance, to the electromagnetic fields in the ether or in ponderable bodies, to fluids or solids as they are treated in hydrodynamics and in the theory of elasticity, to bodies in which a conduction of heat is going on, and to many other cases. We shall only make the assumption that the quantities determining the state of the system are all continuously distributed over space without any abrupt changes. For this purpose we either ignore molecular structure or extend our analysis into the interior of molecules and atoms.

We now fix our attention on an element of volume dV (which we can take to be a parallelepiped $dx\,dy\,dz$) and on the amounts of momentum and energy which it contains. Let G_x, G_y, G_z be the components of momentum and E the energy per unit of volume. Now these quantities can be changed in two ways: namely, by the action of external forces and by actions going on at the surface of the element, that is, interactions between dV and the part of the system surrounding dV.

Interactions of this kind, by which G_x, G_y, G_z can be changed, are called stresses or eventually pressures. We shall as usual denote them by X_x, Y_x, Z_x, X_y etc. The first three refer to an element of surface perpendicular to the axis of x; they represent the components of the force which the matter on the positive side of that element exerts on matter on the negative side. $X_y \cdots$ have similar meanings.

Now if there are no external forces, like gravity, we shall have

$$\frac{\partial G_x}{\partial t} = \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z}. \tag{147}$$

The equation of energy has a similar form,

$$\frac{\partial E}{\partial t} + \frac{\partial S_x}{\partial x} + \frac{\partial S_y}{\partial y} + \frac{\partial S_z}{\partial z} = 0,$$

where S_x, S_y, S_z are the components of the flow of energy. The first term represents the rate of change of E, and the other terms are due to the influence of the actions at the surface of the element. We may say that these terms represent the rate at which work is done by the stresses, but we can just as well say that they are due to a flux of energy across the surface. Indeed, the statements that work is done on a body and that energy is imparted to it are equivalent.

Similarly, the quantities X_x, X_y, X_z can be considered as the fluxes of momentum G_x across elements of surface perpendicular to OX, OY, OZ, or rather as these fluxes with opposite sign. There is one case in which this is the only adequate way of thinking of them; namely, the case in which the actions at an element of surface consist simply in the passage across that element of particles which carry with them their momentum and kinetic energy (kinetic theory of gases).

Four equations of the same form as the above also hold in the system x', y', z', t'. They contain sixteen quantities, X'_x etc., which play the same part in the second system as X_x etc. do in the first. These quantities are related to X_x etc. by transformations comparable to those which we had for g_x, g_y, g_z,

and ϵ. By these formulæ each of the sixteen quantities X'_x etc. is expressed in terms of the corresponding one and a certain number of the other quantities Y_y etc.

How can the transformation formulæ for the quantities now under consideration be derived? Each of them may consist of different parts, — for instance, one due to particles carrying with them momentum and energy, the other of electromagnetic nature. It is clear that a self-consistent system of equations (at least a simple system) can be found only if the formulæ of transformation are the same for the several parts. Assuming this, we have only to find the formulæ for a special case, for which we take the one that has just been mentioned.

Let us suppose that the system consists of material particles all moving with the same velocity v, the number per unit of volume being N. Then

$$G_x = Ng_x, \quad G_y = Ng_y, \quad G_z = Ng_z, \quad E = N\epsilon. \tag{148}$$

What is the amount of energy transferred per unit of time and unit of area across a plane at right angles to OX? The number of particles passing through the plane in unit of time is Nv_x. Thus the flow of energy in the direction OX is

$$S_x = N\epsilon v_x. \tag{149}$$

Similarly, the amounts of momentum (g_x, g_y, g_z) transmitted per unit of time (for which we can write $-X_x$, $-Y_x$, $-Z_x$) are

$$-X_x = Nv_x g_x, \quad -Y_x = Nv_x g_y, \quad -Z_x = Nv_x g_z. \tag{150}$$

These formulæ hold also when v_x is negative, for the current of energy is reckoned as negative when the energy is transferred in the direction of $-x$. Also a momentum g_y transferred in that direction per unit time contributes to $-Y_x$ the amount $-g_y$.

Now, since we have transformation formulæ for the components of v and g and for the energy ϵ, we can easily deduce the transformation formulæ for $G_x \cdots$, $S_x \cdots$, $X_x \cdots$, and E. Only we must take into account the fact that the number N'

of particles per unit of volume in the second system is not equal to N; we have, on the contrary,

$$N' = N\omega = N\left(a - \frac{b}{c}\,v_z\right). \tag{151}$$

This relation follows at once from the previous relation

$$dV' = \frac{dV}{\omega}.$$

In this way we find, for instance,

$$\left.\begin{aligned}
G'_x &= N'g'_x = N\left(a - \frac{b}{c}\,v_z\right)g_x = aG_x + \frac{b}{c}\,X_z \\
S'_x &= N'v'_x\epsilon' = Nv_x(a\epsilon - bcg_z) = aS_x + bcZ_x \\
G'_z &= N'g'_z = N\left(a - \frac{b}{c}\,v_z\right)\left(ag_z - \frac{b}{c}\,\epsilon\right) \\
&= a^2G_z + \frac{ab}{c}\,Z_z - \frac{ab}{c}\,E + \frac{b^2}{c^2}\,S_z \\
E' &= N'\epsilon' = a^2E - \frac{ab}{c}\,S_z - abcG_z - b^2Z_z, \text{ etc.}
\end{aligned}\right\}. \tag{152}$$

We have found these formulæ by considering a simple case; namely, that of a system of particles all having the same velocity. Now suppose that we have particles moving in all directions with very different velocities, as in the kinetic theory of gases. We can decompose such a system into a great number of groups of particles, each group being characterized by a velocity definite in direction and magnitude. The value of each of the quantities $G_x \cdots$ and E will be the sum of parts due to these several groups, and so it is easily seen that the above transformation formulæ will hold for the full values existing in the system, since they hold for the particular values belonging to the individual groups.

We shall now assume that the formulæ (152) hold in all cases, whatever the mechanism to which the stresses etc. are due. We may here remark that in the case of an electromagnetic field in the ether all the quantities occurring in the above formulæ are known; namely, density of energy, flow of energy, electromagnetic momentum, and Maxwell's stresses. We can

express them in terms of the components of **E** and **H** or **E'** and **H'**; and if we introduce the transformation formulæ for **E** and **H**, which have already been given, we can verify (152).

It can also be proved, using (152), that the system of four equations of type

$$\frac{\partial G_x}{\partial t}=\frac{\partial X_x}{\partial x}+\frac{\partial X_y}{\partial y}+\frac{\partial X_z}{\partial z}$$

is equivalent to a set of four equations of the same form in x', y', z', t'.

38. Remarks on Formulæ of Transformation. We have found that there are transformation formulæ for the four quantities dx, dy, dz, dt, the six quantities E_x, E_y, E_z, H_x, H_y, H_z, and the sixteen quantities $X_x \cdots E$. Quantities such as those belonging to these sets, which are grouped together in the transformations, are called components of a "tensor," the nature of the tensor depending on the number of components and the form of the transformation formulæ. If I could spare the time I should be able to discuss the properties of tensors and to give a more elegant form to the above analysis, so that you could almost write down the transformation formulæ from memory, while, with the formulæ as they are now, we can scarcely see any regularity. Later on, when we come to speak of the general theory of relativity (the theory of gravitation), I shall find an opportunity to remedy this. In the equations (148)–(150), which we found for a system of particles all moving with the same velocity, we can of course replace N by the symbol of summation Σ, meaning that the sums are to be taken per unit of volume; that is, the sum for all particles in the element dV is to be divided by dV. In the form which they then take the formulæ will also hold for a system of particles moving in all directions, perhaps in some directions more than in others. Introducing the values of g_x, g_y, g_z, ϵ, we find

$$\left.\begin{array}{lll} X_x=-\sum mkv_x^2, & X_y=-\sum mkv_xv_y, & Y_x=-\sum mkv_yv_x \\ G_x=\sum mkv_x, & S_x=\sum mkc^2v_x & \\ & E=\sum mkc^2 & \end{array}\right\}. \quad (153)$$

Notice that it is not even necessary that m should be the same for all particles (mixture of gases).

Some of these formulæ are well known in the theory of gases. If the gas molecules move equally in all directions, we have

$$X_x = -\tfrac{1}{3}\sum mkv^2,$$

and if we neglect the difference between k and unity,

$$X_x = -\tfrac{1}{3}\sum mv^2,$$

meaning that there is a pressure

$$\tfrac{1}{3}\sum mv^2.$$

The tangential stresses X_y etc. appear in the theory of viscosity. We find, for any values of the velocity and of k,

$$X_y = Y_x \text{ etc.} \tag{154}$$

and $$S_x = c^2G_x, \quad S_y = c^2G_y, \quad S_z = c^2G_z, \tag{155}$$

or, in vector form, $$\mathrm{S} = c^2\mathrm{G}. \tag{156}$$

These relations (154) and (156) also hold for an electromagnetic field in the ether, and we shall suppose them to be generally true. The equations (154) have been known for a long time in the theory of elasticity. The transformation formulæ show that if (154) and (156) hold in the system x, y, z, t, they also hold in the system x', y', z', t'. For instance, if $Z_x = X_z$ and $S_x = c^2G_x$, we see by the first two of the equations (152) that $S'_x = c^2G'_x$.

The relation (156) between momentum and flow of energy is very remarkable. It was first found in the theory of electromagnetism, and it seems at first sight strange that there should be this relation between quantities that have wholly different meanings. That the relation exists is due to the circumstance that for material points we wrote, for the energy,

$$\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} = mc^2 + \frac{1}{2}mv^2 + \cdots, \tag{157}$$

including the term mc^2.

Suppose, for instance, that particles all move in the direction of OZ with a velocity v which is so small that squares of v can

be neglected. Then there is a momentum Nmv per unit volume. How can there be a current of energy c^2 times as great? Simply because each particle has the energy mc^2 and carries it along, giving a total current of energy Nmc^2v.

If we think only, as we did in former days, of ordinary kinetic energy, $\frac{1}{2}mv^2$, there will be (if terms in v^2 are neglected) neither an energy nor a flow of energy.

The inverse case can likewise occur; that is, there are cases in which there is a current of energy, while we do not, at first sight, recognize that there is a momentum.

Consider, for instance, a vertical column of a gas consisting of equal molecules and inclosed in a cylindrical vessel. The upper and lower ends of the vessel are maintained at different constant temperatures, the higher temperature being at the top. A stationary state will be attained, in which the gas as a whole is at rest, and as we are accustomed to think of a moving body when speaking of momentum (quantity of motion), we might think there was no momentum; yet there is a downward flow of energy, the conduction of heat.

The solution of this apparent paradox is very easy if we use our formulæ. Let us choose the axis of z vertically downward. The number of particles which, per unit of area, cross a horizontal section in a downward direction exceeds the number of particles which cross the same section in an upward direction by the quantity Σv_z per unit of time, the sum being extended over unit of volume. Therefore in the stationary state we have the relation

$$\sum mv_z = 0.$$

If, then, the momentum of a particle were mv_z, we should have no resulting momentum, but according to the principle of relativity the momentum is

$$\sum \frac{mv_z}{\sqrt{1-\frac{v^2}{c^2}}},$$

and this can very well be different from zero, though $\Sigma mv_z = 0$. This is just what occurs when there is a temperature gradient. The particles going downward will have greater velocities than

those that go upward. Therefore for the positive values of v the expression

$$\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}$$

will be greater than for the negative values. Now if you have a sum $\Sigma m v_z$ that is zero, and if you then multiply the positive terms by factors greater than those by which you multiply the negative terms, the sum will become positive.

Similarly, a current of heat in a metal must be accompanied by a momentum, but a calculation shows that the latter is much too small to be observed. It will be likewise impossible to prove experimentally that in a crystal the electromagnetic momentum of a beam of light is in the direction of the rays and not normal to the wave front; one may expect this on theoretical grounds because Poynting's vector is in the direction of the ray. I have insisted on all this because both (154) and (156) are essential in Einstein's theory of gravitation. If they were not satisfied, the general equations by which he determines the gravitational field would be self-contradictory.

The way in which the equations can be verified will, as in our example from the theory of gases, depend on our conception of the mechanism. I can say no more about it than this: if we were able to analyze phenomena in full detail, and to fix our attention on values of G and S in the interior of atoms and even of electrons, relativity requires that then we should find $S = c^2G$. It can well be understood that if a very close scrutiny revealed the existence of the relation at all points, its validity in a more superficial inspection would then follow.

39. Momentum and Energy of a System of Any Kind in a Stationary State. By means of the transformation formulæ for stresses, energy, etc. we can prove that the formulæ by which we defined momentum and energy have a much wider application than we expected when establishing them. They apply to any system that is in a stationary state; that is, a system that does not in the long run get farther and farther away from its initial state, though it may be the seat of internal

motions or changes. It may be an electron, an atom, a molecule, a cavity filled with black radiation, — any body, even a planet or the sun itself.

The question is, however, What are we to understand by v? We cannot speak of the center of gravity of a system, but we can reason as follows:

In the first system of coördinates (system 1) the body will have a certain resultant momentum whose value we can determine when we have sufficient knowledge of the body's state. We can suppose, without loss of generality, that this resultant momentum has the direction OZ.

Let us now pass on to a second system of coördinates (system 2), using a relativity transformation with constants a and b. We can then calculate, by means of the transformation formulæ, the resultant momentum in this system 2. Now the constants a and b can be chosen in such a way that in system 2 the resultant momentum is zero. It is then natural to say that the body as a whole is at rest in system 2, and that in system 1 it has the same velocity as the origin of system 2 has in system 1, — a velocity which we know because we know a and b.

This is our definition of v; if we follow it we can always use the expressions

$$\frac{mv}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{and} \quad \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

for momentum and energy.

Proof. Suppose that in the system x, y, z, t there is no resultant momentum but a certain amount of energy mc^2. By means of this quantity of energy we can define a coefficient m, the mass of the system. If we calculate the integrals for a definite value of the time t, we have

$$\int G_z\,dV = 0, \quad \int E\,dV = mc^2.$$

From these we can derive the value of $\int G'_z\,dV'$ taken for a definite value τ of the time t', as we must of course do when we want to know the momentum of the body in the second system.

By virtue of (152), and taking into account the relation $S = c^2 G$, we have

$$G'_z = (a^2 + b^2)G_z + \frac{ab}{c} Z_z - \frac{ab}{c} E.$$

Let the expression on the right-hand side be denoted by $\phi(x, y, z, t)$; then we have

$$\int G'_z \, dV' = \frac{1}{a} \int \phi\left(x, y, z, \frac{\tau}{a} + \frac{bz}{ac}\right) dV,$$

since for a definite value τ of t', $t = \frac{\tau}{a} + \frac{bz}{ac}$ and, by virtue of (121),

$$dx' = dx, \quad dy' = dy, \quad dz' = \frac{1}{a} dz, \quad \text{giving} \quad dV' = \frac{1}{a} dV.$$

We know beforehand that the result for the momentum will be a constant; we can therefore replace ϕ by its mean value

$$\overline{\phi}\left(x, y, z, \frac{\tau}{a} + \frac{bz}{ac}\right)$$

over all values of τ. But when, for definite values of x, y, z, τ varies over a long interval of time, $\frac{\tau}{a} + \frac{bz}{ac}$ does so likewise. We may therefore write, instead of the above mean value, the mean value $\overline{\phi}(x, y, z, t)$ taken over a long interval of t. Thus

$$\int G'_z \, dV' = \left(a + \frac{b^2}{a}\right) \int \overline{G}_z \, dV + \frac{b}{c} \int \overline{Z}_z \, dV - \frac{b}{c} \int \overline{E} \, dV.$$

Here the first integral is zero, as we have assumed. The second is zero in any stationary state.* Hence, by virtue of

$$\int \overline{E} \, dV = mc^2,$$

we have

$$\int G'_z \, dV' = -bmc.$$

* *Proof.* Let $\int Z_z \, dx \, dy = Q$ for any definite value of z, the integral extending over a plane perpendicular to OZ. Then, on account of the meaning of Z_z and G_z,

$$\frac{dQ}{dz} = \frac{d}{dt} \int G_z \, dx \, dy.$$

Thus on the average over a long time $\frac{d\overline{Q}}{dz} = 0$ and $\overline{Q} = 0$, because, for $z = -\infty$, $\overline{Q} = 0$. But

$$\int \overline{Z}_z \, dV = \int \overline{Q} \, dz.$$

Hence the integral on the left-hand side is zero.

Similarly, $$\int E'dV' = amc^2.$$

Now the origin of the system of coördinates x, y, z, t (in which our body has no resultant motion) has in the system x', y', z', t' a velocity

$$v = -\frac{b}{a}c$$

in the direction OZ; and since, therefore,

$$a = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \quad b = -\frac{v}{c\sqrt{1-\frac{v^2}{c^2}}},$$

the above expressions become

$$\frac{mv}{\sqrt{1-\frac{v^2}{c^2}}} \text{ and } \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}.$$

40. Relativity and the Principles of the Conservation of Energy and Momentum. In the expressions for momentum and energy m is constant, whatever be the system of coördinates. The formula for momentum has been derived in the first place for electrons and has been verified by experiments. It was thought at the time to prove that the mass of an electron is entirely electromagnetic. We must now say that the formula is general and quite independent of any conception of the nature of mass. We can also be sure that it will apply to other models of electrons that have been or may be imagined, also to models of atoms. Let us now consider some applications of our formulæ.*

1. As a first example we take the impact of two equal inelastic bodies. Heat is developed, and we shall suppose it not to be radiated but to remain in the bodies.

Suppose in the first place that, for an observer A, the first body has a velocity $v_z = +q$, the second a velocity $v_z = -q$. After impact both have velocity zero. Of course the law of the conservation of momentum holds. The energy is also conserved because the mass has changed (the bodies having attained new states).

* See also Note 10, Appendix.

Let the initial mass of each be m and the final one $\overline{m}$; then

$$\frac{mc^2}{\sqrt{1-\frac{q^2}{c^2}}} = \overline{m}c^2,$$

or

$$\overline{m} = \frac{m}{\sqrt{1-\frac{q^2}{c^2}}}.$$

Relative to an observer B who uses the x', y', z', t' system of coördinates with

$$a = \frac{1}{\sqrt{1-\frac{q^2}{c^2}}}, \quad b = \frac{q}{c\sqrt{1-\frac{q^2}{c^2}}},$$

$$v_z' = \frac{av_z - bc}{a - \frac{b}{c}v_z} = \frac{v_z - q}{1 - \frac{qv_z}{c^2}},$$

the velocities before collision are 0 and

$$-\frac{2q}{1+\frac{q^2}{c^2}}$$

respectively, while after collision they are both $-q$. Before collision the momenta are 0 and

$$-\frac{2mq}{1-\frac{q^2}{c^2}}$$

respectively, while after impact they are

$$-\frac{\overline{m}q}{\sqrt{1-\frac{q^2}{c^2}}} = -\frac{mq}{1-\frac{q^2}{c^2}},$$

for each body. The energies before impact are mc^2 and

$$mc^2\frac{1+\frac{q^2}{c^2}}{1-\frac{q^2}{c^2}},$$

while after impact each body has the energy

$$\frac{\bar{m}c^2}{\sqrt{1-\frac{q^2}{c^2}}}=\frac{mc^2}{1-\frac{q^2}{c^2}}.$$

For each observer the laws of conservation are confirmed.

The total momentum was 0 for A and

$$-\frac{2\,mq}{1-\frac{q^2}{c^2}}$$

for B, while the total energy was

$$\frac{2\,mc^2}{\sqrt{1-\frac{q^2}{c^2}}}$$

for A and

$$\frac{2\,mc^2}{1-\frac{q^2}{c^2}}$$

for B.

2. Suppose that a system divides itself into two parts with masses m_1 and m_2, moving in the direction OZ with velocities v_1 and $-v_2$. If the original system had a mass m and a velocity 0, we have

$$\frac{m_1v_1}{\sqrt{1-\frac{v_1^2}{c^2}}}-\frac{m_2v_2}{\sqrt{1-\frac{v_2^2}{c^2}}}=0,$$

$$\frac{m_1}{\sqrt{1-\frac{v_1^2}{c^2}}}+\frac{m_2}{\sqrt{1-\frac{v_2^2}{c^2}}}=m.$$

By means of these formulæ we can, for instance, calculate m_1 and v_1 when m, m_2, and v_2 are known.

Suppose that m_2 is very small compared with m (for example, an α particle shot off by an atom). Then to a first approximation

$$m_1=m-\frac{m_2}{\sqrt{1-\frac{v_2^2}{c^2}}},\quad v_1=\frac{m_2}{m_1}\frac{v_2}{\sqrt{1-\frac{v_2^2}{c^2}}}. \tag{158}$$

This shows that if an atom loses an α particle, the mass of the atom is diminished by somewhat more than the mass of that particle; namely, by its energy divided by c^2. In general we may draw the following conclusion:

If a system which is originally at rest in our system of coördinates is again so after the lapse of a certain time, but in a changed state, then, if the energy is different from what it was before, we shall have to attribute a new value to the mass; for m was defined by the equation

$$\text{Energy} = mc^2.$$

In particular the mass of the sun must be supposed to change on account of the loss of energy by radiation.

REMARK 1. The two observers A and B will not ascribe the same energy to a system. Therefore we cannot, without making some statement about the system of coördinates, say that the universe has a definite energy; but for each of the observers the law of the conservation of energy will hold.

REMARK 2. We spoke of the changes which the principle of relativity has obliged us to make in the principles of mechanics, so that we now take as the equations of motion of a particle

$$\frac{d}{dt}\left(\frac{mv_x}{\sqrt{1-\frac{v^2}{c^2}}}\right) = F_x, \text{ etc.} \tag{159}$$

Of course, before we can do anything with these equations we shall have to find out what are the forces that act on the particle; that is, what functions of the relative coördinates and perhaps velocities we shall have to introduce in (159). Now it is clear that the principle of relativity can hold only when the forces F in the first system and the forces F' in the second system depend in the same way on the coördinates and velocities. So the principle implies definite assumptions regarding the law of force.

REMARK 3. We cannot work this out, but one thing can be remarked; namely, that the principle requires that all forces should be transmitted with the velocity of light.

Suppose that a particle at O acts in some way on another particle at P. At time $t = 0$ a sudden change is produced at O; for instance, the particle may be suddenly shifted. When will P feel the result

of this change? The effect cannot be instantaneous, for if it were instantaneous in the x, y, z, t system it would not be so in the x', y', z', t' system. This is contrary to the principle.

If, on the other hand, we assume that the action is propagated with velocity c, this will hold in both systems.

41. Experimental Verifications of the Principle of Relativity. These may be classified as follows:

1. Absence of effects due to the motion of the earth.
2. Experiments on the propagation of light in moving bodies.
3. Experiments on the deflection of moving electrons (cathode rays and β rays) in electric and magnetic fields, by which, for the case of the electron, the new formula for the momentum has been confirmed.
4. The explanation given by Sommerfeld of the fine structure of spectral lines and the extension of that explanation to Röntgen spectra.

The principle of relativity leads to many further important and interesting questions, but it will be better to discuss these in connection with the theory of gravitation. For the present we turn to some brief considerations relating to the structure of matter.

42. Structure of the Electron. The formula for momentum was found by a theory in which it was supposed that in the case of the electron the momentum is determined wholly by that of the electromagnetic field; namely,

$$\frac{1}{c}\,[EH] \text{ per unit of volume.}$$

This meant that the whole mass of an electron was supposed to be of electromagnetic nature. Then, when the formula for momentum was verified by experiment, it was thought at first that it was thereby proved that electrons have no "material mass." Now we can no longer say this. Indeed, the formula for momentum is a general consequence of the principle of relativity, and a verification of that formula is a verification of the principle and tells us nothing about the nature of mass or of the structure of the electron. Therefore physicists are

absolutely free to form any hypotheses on the properties and size of electrons that may best suit them. You can, for instance, choose the old electron (a small sphere with charge uniformly distributed over the surface) or Parson's ring-shaped electron,* endowed with rotation and therefore with a magnetic field; you can also make different hypotheses about the size of the electron. In this connection I may mention that A. H. Compton's experiments on the scattering of γ rays by electrons † have led him to ascribe to the electron a size considerably greater than it was formerly supposed to have.

Of course I need hardly mention that, whatever theory we favor, we must suppose that a motion of translation will make the electron contract. Indeed, we want to apply the principle of relativity to the electron also; if then we know what is going on in the electron when it has no motion of translation, we can deduce from the principle in full detail the state that will exist when there is such a motion.

The quantities which we know (for an electron without translation) are the charge e and the energy ϵ. This leads us to ascribe to the electron a mass ϵ/c^2. Now, if ϵ were wholly electrostatic energy due to some distribution of the charge e, the dimension (diameter) of the electron ought to be of the order of magnitude e^2/ϵ, if we express the charge e in the usual electrostatic units. Indeed, the electrostatic energy is e^2/r, where r is some mean distance. If we want to have the electron much larger, we must assume that only a part, perhaps a small part, of the total energy is electrostatic energy.

We have already been led to the assumption that in an electron there are forces different from electromagnetic forces and an energy corresponding to these forces. If we consider the charge to be at rest when the electron has no translation, we must necessarily introduce forces which can counterbalance

* A. L. Parson, A Magneton Theory of the Structure of the Atom. Smithsonian Publication No. 2371, Washington, 1915.

† A. H. Compton, *Phys. Review*, Vol. 14 (1919), p. 20. Compton ascribed a radius of $2 \cdot 10^{-10}$ centimeter to the ring electron, while the radius generally adopted for the old electron was about 10^{-13} centimeter. Parson estimated the radius of the ring electron to be $1.5 \cdot 10^{-9}$ centimeter.

the electrostatic repulsions and may prevent them from making the electron explode; or, what amounts to the same thing, we have to balance by some other forces the stresses in the surrounding electric field, which would tear the parts of the electron asunder. In Poincaré's model of the electron,* there is a sphere over which the charge is uniformly distributed and in the interior of which there is a tension (though there is no electric field) equal to the Maxwell's stress that is exerted by the surrounding field. Corresponding to this tension there is a certain amount of accessory energy (equal to the third part of the electric energy) existing in a space where there is no electric field; this can scarcely be called an electromagnetic energy, though of course it may be very intimately connected with the electromagnetic phenomena.

43. The Structure of Atoms. According to a hypothesis which is due to Rutherford and which has been further developed by Bohr, — a hypothesis that has proved to be of the utmost value, — an atom consists of a positively charged nucleus and a certain number of electrons revolving around it.

If we take as a natural unit of charge the charge of a monovalent positive or negative ion (you know that the charge of the electron is such a unit), the charge of the nucleus is given by the "atomic number" of the element. If the atom is in its neutral state, so that the total charge is zero, the atomic number N also gives us the number of electrons surrounding the nucleus; if there are more or less of them, we have a charged atom or an ion.

The electrons are supposed to move under the attraction of the nucleus and their mutual repulsions, according to the laws of mechanics (eventually as in Sommerfeld's theory of the fine structure with the modifications required by relativity); according to Bohr, however, not all states of motion that would be possible according to these laws occur in reality, but only certain select ones that Bohr calls the "stationary states."

* H. Poincaré, "Sur la dynamique de l'électron," *Rend. Palermo*, t. 21 (1906), p. 129. See also H. A. Lorentz, The Theory of Electrons, p. 213; Th. de Donder, La gravifique einsteinienne, pp. 95, 136.

They are determined by conditions in the mathematical expression of which you always find Planck's constant h, and which are properly called quantum conditions.

Let me now call your attention to a beautiful theorem which has often been used in the kinetic theory of gases and which may also be applied to an atom; it is the theorem of the so-called virial.* In formulating it we shall neglect the relativity terms.

Consider a system of material points acted on by forces X, Y, Z. For each of them we have the equations of motion

$$X = m\frac{d^2x}{dt^2}, \quad Y = m\frac{d^2y}{dt^2}, \quad Z = m\frac{d^2z}{dt^2}. \tag{160}$$

We multiply these equations by the coördinates of the point considered and form the equation

$$\sum(Xx + Yy + Zz) = \sum m\left(x\frac{d^2x}{dt^2} + y\frac{d^2y}{dt^2} + z\frac{d^2z}{dt^2}\right). \tag{161}$$

For the second member we can write

$$\frac{d}{dt}\left[\sum m\left(x\frac{dx}{dt} + y\frac{dy}{dt} + z\frac{dz}{dt}\right)\right] - \sum m\,(\dot{x}^2 + \dot{y}^2 + \dot{z}^2), \tag{162}$$

where the last term is double the kinetic energy T of the system. Now take mean values over a long interval of time. If the system is in a stationary state, the mean value of any quantity that is the derivative with respect to the time of a quantity determined by the state of the system will vanish.†

* A form of the virial theorem was used by Jacobi, *Crelle's Journal*, Bd. 17, p. 81; Bd. 28 (1845), p. 276; *Vorlesungen über Dynamik* (1842–1843), p. 22. The name *virial* seems to have been introduced by Clausius when he applied the theorem in a general way to calculate the relation between pressure, volume, and temperature in a perfect gas (*Ann. Phys. Chem.*, Bd. 141 (1870), p. 124; *Jubelband* (1874), p. 411; *Phil. Mag.* (Aug., 1870)). The virial theorem has been used recently by Poincaré and Eddington in their astronomical work. See Jeans, Problems of Cosmogony and Stellar Dynamics, chap. viii.

† The mean value of $\frac{d\phi}{dt}$ is, in fact,

$$\frac{1}{t - t_0}\int_{t_0}^{t}\frac{d\phi}{dt}\,dt = \frac{\phi(t) - \phi(t_0)}{t - t_0}.$$

The numerator is finite, but the denominator can be made to exceed any limit.

Therefore, indicating mean values by bars

$$\sum \overline{(Xx + Yy + Zz)} = -2\,\overline{T}. \tag{163}$$

The expression on the left-hand side (the virial of the forces) takes a simpler form when only electrostatic attractions and repulsions occur. Consider two points x, y, z and x', y', z' at a distance r apart and exerting on one another a force k/r^2, where k is positive for a repulsion. Then the components of the force acting on the first point are

$$k\,\frac{x - x'}{r^3} \text{ etc.,}$$

and those of the force on the second point are

$$-\,k\,\frac{x - x'}{r^3} \text{ etc.}$$

Thus the part which these forces contribute to the virial is

$$k\,\frac{(x - x')^2 + (y - y')^2 + (z - z')^2}{r^3} = \frac{k}{r},$$

so that we get

$$\overline{\sum \frac{k}{r}} = -2\,\overline{T}. \tag{164}$$

But the sum on the left-hand side is exactly the potential energy of the system if we take the potential energy to be zero when the particles are at an infinite distance. Therefore, if U is the potential energy,

$$\overline{U} = -2\,\overline{T}. \tag{165}$$

This formula is easily verified in the case of circular motion, but it holds in all cases. It shows, for instance, that as soon as one of the two energies is determined by a quantum condition, the same is true of the other energy.

44. Applications of the Virial Theorem. 1. When the particles are removed from each other to an infinite distance in such a way that after the separation they are at rest, the energy is increased by

$$-\,(\overline{U + T}) = \overline{T}.$$

Hence the mass of the system is smaller than the sum of the masses of the isolated constitutive particles, the difference being

$$\frac{\overline{T}}{c^2} = \frac{1}{2\,c^2}\sum mv^2.$$

2. Relation (165) shows that $\overline{U}$ must be negative; in the sum $\sum \frac{k}{r}$ the terms corresponding to an attraction must predominate. This is easily understood, because if the particles are to remain together, notwithstanding their motion, they must attract each other. We can also write

$$\overline{\sum \frac{ee'}{r}} = -2\,\overline{T}$$

and conclude that the terms in which e and e' have opposite signs (attraction) must predominate over those in which the signs are the same.

3. We can apply this result, though with much diffidence, to the constitution of the nucleus. Thanks to the splendid results obtained by Rutherford and others, we now know something about it. It is highly probable that the nucleus of the higher elements is a very complicated structure, quite a microcosm, containing smaller positive nuclei (those of helium, for instance) and a certain number of electrons; if the element is radioactive, both the positive nuclei of helium and the negative electrons can be expelled in the form of α-rays and β-rays.

It is clear that if the nucleus is to be built up by means of electric forces, the electrons within it are absolutely necessary as binding material, and that we ought to have a sufficient number of them. Now there is a difficulty because the number which we can really suppose to exist in the nucleus is in many cases rather small; this is due to the fact that the nucleus as a whole must have a positive charge.

In his Bakerian lecture on the Nuclear Constitution of Atoms Professor Rutherford has given us a picture of the atom of oxygen (Fig. 33). The numbers 3, 4 give the masses of the five positive particles that are supposed to exist; each of them has a charge 2, as is indicated by the double signs $++$.

In addition to these five particles there are two electrons whose mass can be neglected and which are simply represented by the signs —, indicating their charges.

You see that the total mass is 16 and the total charge 8, as they ought to be, since the atomic weight of oxygen is 16 and the atomic number 8. The question now arises, Can the 10 negative terms in the virial due to the action of the electrons on positive charges predominate over the 11 positive terms which represent the repulsions between charges of the same sign, though the positive numerators ee' are greater than the negative ones? As a matter of fact, the discussion is of little importance, for it seems very improbable that the ordinary laws of mechanics and of electrostatics will apply to the interior of the nucleus. There may be very rapid motions, and under the influence of these and of the mutual interactions the particles may be deformed. There may also be a magnetic field due to the motions, which also produces certain forces acting on the electrons.

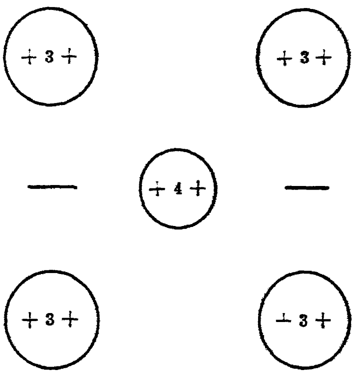

FIG. 33

Finally, there may be forces which are not electromagnetic, or there may be a deviation from Coulomb's law at very short distances; of course such a deviation can be obtained by superposing some new field of force on the ordinary electrostatic field.

It is rather an interesting fact that the theorem of the virial can be extended so as to cover all these complications. As the velocities may be very great, I shall remark in the first place that our equation for the mean value of the virial V in the case of a system of material particles

$$\overline{V} = -2\,\overline{T}$$

must be somewhat changed when we introduce the mechanics of relativity. We can then write *

$$\overline{V} = -2\,\overline{T}', \tag{166}$$

where

$$T' = \frac{1}{2}\sum \frac{mv^2}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{167}$$

Let us further consider a system of electric charges moving in any way and producing an electromagnetic field. We shall suppose (for the sake of generality) that these charges are associated with certain masses of non-electromagnetic origin. Let the volume density of these masses be μ and let T' be the quantity for them that corresponds to T' in the above case; namely,

$$T' = \frac{1}{2}\int \frac{\mu v^2}{\sqrt{1-\frac{v^2}{c^2}}}\,dS. \tag{168}$$

We shall suppose that certain forces of non-electromagnetic origin act on these charged masses, and we shall further suppose that in the electromagnetic field produced by the charges there is nothing that is not taken into account in Maxwell's equations.

Let V be the virial of the forces just mentioned and let E be the total electromagnetic energy of the system. Then†

$$\overline{V} = -2\,\overline{T}' - \overline{E}. \tag{169}$$

Example 1. The virial of a uniform normal pressure p distributed over the boundary surface of a volume S is $-3\,pS$. Hence if a gas is inclosed in the volume S,

$$-3\,pS = -2\,T',$$

or

$$p = \frac{2}{3}\frac{T'}{S}. \tag{170}$$

For the case $T' = T$ this is the fundamental formula of the kinetic theory of gases.

* See Note 11, Appendix. † See Note 12, Appendix.

EXAMPLE 2. Take a vessel filled with black radiation. On account of the radiation pressure we must apply an external pressure p. Now we have, by virtue of (169),

$$-3\,pS = -E,$$

or

$$p = \frac{1}{3}\frac{E}{S}. \tag{171}$$

This is the ordinary formula for radiation pressure.

EXAMPLE 3. Consider Poincaré's electron of radius R with a charge e distributed over the spherical surface, the electron being at rest. Poincaré's tension is equal to Maxwell's stress on the outside surface. In rational units it is

$$p = \frac{e^2}{32\,\pi^2 R^4}.$$

The electrical energy in the outside space is $\frac{e^2}{8\,\pi R}$, and relation (169),

$$-3\,p\cdot\frac{4}{3}\,\pi R^3 = -\frac{e^2}{8\,\pi R},$$

is verified.

The general formula (169) may well apply to the nucleus. There must be certain forces of non-electromagnetic origin. As both T' and E are essentially positive, the virial of these forces must be negative, and this shows that these forces are necessary to hold the charge together. They might, for instance, be confined to the interior of the charged particles (like Poincaré's tension), or they might act between one charged particle and another.

However this may be, it is certain that in Bohr's model of the atom there is scarcely any room left for non-electromagnetic forces. At least in those cases for which the theory has been fully developed we can say that between the nucleus and the electron revolving around it there are only electric forces. It is therefore natural to think that between a nucleus and more distant electrons there are also no other forces. If we extend this to the mutual action between nuclei and that between electrons we are led to the idea that the forces of cohesion and those that manifest themselves in cases of elasticity also have a purely electromagnetic origin.

45. How far can Molecular Forces be explained by Electromagnetic Actions? Some interesting investigations connected with this question have been made by Born and Landé,* and I should like to say a few words about them, confining myself to the case of a crystal like that of sodium chloride. You know that there cannot be any doubt that in such a crystal atoms of Na and Cl are arranged at the points of a cubic space lattice, the atoms of both kinds alternating regularly along each line of the lattice that has the direction of an edge of the elementary cube. I mean to say that the nuclei of the Na and the Cl atoms are arranged in this way, while around these nuclei the electrons are revolving in certain orbits which we can suppose to be determined by quantum conditions. As the atomic number of Na is 11 and that of Cl is 17, the nuclei have charges $+11\,e$ and $+17\,e$, where $-e$ is the electronic charge. The atoms are supposed to be ionized so that there are 10 electrons revolving around each Na nucleus and 18 around each Cl nucleus.

Born and Landé make definite assumptions about the arrangement and motion of these electrons. They suppose them to be arranged in rings, and they even replace the electrons that follow one another on the same path by a uniform distribution of electric charge around the ring. Having done this they speak no longer of a motion of the rings. Moreover, they give to these rings definite magnitudes and positions. The magnitudes are determined by quantum conditions and are supposed not to be changed by the interaction between neighboring atoms. As to the position, or rather orientation of the rings, it is chosen in such a way that it is in agreement with the cubical symmetry of the crystal.

You see that all these assumptions amount to this: we suppose the points of our lattice to be the centers of small, rigid bodies having definite and invariable distributions of electric charges. We want to explain the properties of the crystal by means of the electric actions between these bodies.

* *Berlin. Sitzungsber.* (1918), p. 1048; *Verh. d. D. Phys. Ges.*, Bd. 20 (1918). pp. 202, 210; *Ann. d. Phys.*, Bd. 61 (1920), p. 87. M. Born–E. Bormann, *Verh. d. D. Phys. Ges.*, Bd. 21 (1919), p. 733; *Ann. d. Phys.*, Bd. 62 (1920), p. 218.

It will be convenient to decompose the distribution of charges, with which we are concerned, in a particular way. The sodium atom is supposed to have lost one electron. We may therefore suppose that in the sodium particle we have a charge $+e$ in the nucleus, which we shall call the residual charge, and in addition to this a charge $+10e$ in the nucleus and $-10e$ in the revolving electrons. These $+10e$ and $-10e$ taken together form what we may call a neutral atom. Similarly, the chlorine atom has one electron more than it has in its natural state. We can imagine its constitution as follows: a charge $-e$ in the nucleus (the residual charge), also a second charge $+18e$ in the nucleus and a compensating charge $-18e$ in the revolving electrons, the two latter charges forming a neutral atom. We thus get a residual charge and a neutral atom at each point of the lattice.

The theory of the configuration which a system of particles will assume under the action of the mutual forces, the problem being regarded as a statical one, can be based on a consideration of the potential energy U of a cubical distribution, of the type considered, for different lengths l of the edge of an elementary cube, say

$$U = F(l).$$

If we leave the body to its internal actions, the particles will assume positions for which U is a minimum. This configuration is determined by

$$F'(l) = 0,$$

an equation which will give us the length $l = l_0$ in the state of equilibrium. If we further consider the change which is produced in the potential energy by a small change in l, that is, by a uniform expansion or contraction, omitting terms of an order higher than the second, we shall find the coefficient of compressibility.

Born and Landé have calculated the function $F(l)$ and have deduced from it the distance l_0 between neighboring atoms, not only in the case of NaCl but also for some other regular crystals. In order to see clearly the basis of this calculation we must remark that by their theory the authors find the ratio between

l_0 and the diameter of the rings. These diameters have been determined by quantum conditions, and so the mutual distances between the atoms of the solid body, and consequently the density of the body, are seen finally to depend on Planck's constant h.

The calculated values of $2l_0$ in terms of a unit equal to 10^{-8} centimeters are given in the following table and are compared with values derived from the known values of the density and the number of molecules in a gram molecule (Avogadro's constant).

LiF	LiCl	NaF	NaCl	KF	KCl
3.60	4.19	4.86	5.44	5.34	5.31
4.00	5.11	4.60	5.59	5.90	6.24

Born and Landé have also calculated values for the compressibility that agree satisfactorily with experimental data. I shall not reproduce the calculations of l_0 and of the compressibility but shall simply say that the potential energy was found to have the form

$$U = -\frac{\alpha}{l} + \frac{\beta}{l^5},$$

where α and β are positive constants. To determine l_0 we have the equation

$$\frac{dU}{dl} = \frac{\alpha}{l^2} - 5\frac{\beta}{l^6} = 0,$$

giving

$$l_0 = \left(\frac{5\beta}{\alpha}\right)^{\frac{1}{4}}.$$

You see that the possibility of determining l_0 is due to the fact that in U there are two terms which depend on l in different ways, and whose derivatives have opposite signs. These two terms remind us of the two kinds of molecular forces to which physicists had recourse in former days in the consideration of problems of this kind. The atoms or molecules were supposed to act on one another in two ways, — namely, with attracting and with repulsive forces, — and these were regarded as counterbalancing each other in the state of equilibrium.

A term like $-\frac{\alpha}{l}$ can be considered as due to attractions, and a term like $\frac{\beta}{l^5}$ as due to repulsions. Of course, if the equilibrium is to be stable, the repulsive forces must increase, when the particles are made to approach each other, to a higher degree than the attractive forces. This is the case when the exponent of l in the denominator is higher for them than for the attractive forces. (On account of this the value of $\frac{d^2U}{dl^2}$ for $l = l_0$ is positive, so that U is really a minimum.)

Now the first part of the above expression for U is due to the action between what I have called the residual charges, whereas the second term is due to the forces between these charges and the neutral atoms, and also to the forces between the neutral atoms themselves. We can therefore say that in Born and Landé's theory the action between the residual charges takes the place of the old attractions, and the action exerted by the neutral atoms that of the old repulsions. Indeed, we can see that the residual charges give rise to attractions, for the residual charges nearest to a given one have always the opposite sign. As to the neutral atoms, we may remark that two atoms A and B turn toward each other their negatively charged rings, and the neighboring parts C and D of these will repel each other (Fig. 34). You can even see that if AB is diminished in a slight degree this repulsion will increase more rapidly than the attraction, for since AC and BD do not change, CD will diminish just as much as AB; but this same decrease will be a greater part of CD than it is of AB, and the change in the force depends on the ratio in which the distance is diminished. This reasoning, however, like the formulæ, is based on the assumption that the dimensions of the rings do not change with the distance AB, — an assumption that ought, of course, to be justified.

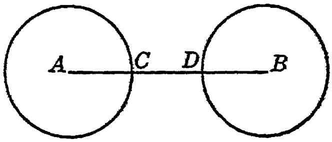

FIG. 34

The agreement between the theoretical results and what we know regarding the atomic distances and the compressibility is

so satisfactory that we cannot help thinking that there must be a good deal of truth in the theory. But there are also serious difficulties which I must point out to you. In the first place, it can be shown that a system of charged particles can never be in stable equilibrium under the influence of these mutual attractions and repulsions alone. This is an extension of the theorem of Earnshaw which we have already had occasion to use.

Of course stationary states of motion, and not of rest, can very well be stable, and if they were not so according to ordinary mechanics, the stability could perhaps be insured by properly chosen quantum conditions; it is clear that we can maintain stability if we simply exclude as nonexistent those perturbations of an unstable state which would become disastrous to the system.

But in Born and Landé's theory there is no question of all this; they simply consider the crystal as built up of rigid bodies bearing electric charges, and treat the problem as one of pure statics. Then their equilibrium must certainly be unstable, and this will show itself by a negative value of one of the moduli of elasticity, which of course is inadmissible. A calculation shows that this is really the case.*

In the second place, I must make some remarks about the second term in the formula for the potential energy. The occurrence in this term of a high power of l in the denominator is due to the fact that the term arises from the action exerted by neutral atoms. We are led already to a term varying inversely as the square of the distance when we consider the potential due to a doublet consisting of two equal and opposite charges at a short distance from one another. If, for instance, there is a charge $-e$ at the point $P(x, y, z)$ and $+e$ at the point $x+\delta, y, z$, the potential at a point x', y', z' at distance r from P will be (when δ is very short)

$$\phi = e\delta \frac{\partial}{\partial x}\left(\frac{1}{r}\right) = e\delta \frac{x'-x}{r} \cdot \frac{1}{r^2}.$$

* See Note 13, Appendix.

In other cases the rapidity with which the potential increases will be still greater. We can, for instance, imagine a doublet, like the one we considered just now, at the point x, $y + \epsilon$, z, and a second doublet, equal and opposite to it, at the point x, y, z, ϵ again being infinitely short. Then the potential at P' will be

$$\epsilon \frac{\partial \phi}{\partial y} = e\delta\epsilon \frac{\partial^2}{\partial x \, \partial y}\left(\frac{1}{r}\right) = 3\, e\delta\epsilon \frac{(x'-x)(y'-y)}{r^2} \cdot \frac{1}{r^3}.$$

We can say that the more the actions of the charges in a neutral system destroy one another at external points, the higher will be the power of r in the denominator. So the factor l^5 in Born's formula is to be explained. Born has even imagined a distribution of electrons for which the exponent will be 9 instead of 5, and by which he can get a still better agreement with some measured data.

This distribution, however, is somewhat different from that to which he had first been led, following Bohr's views. The mutual potential energy of two neutral atoms will, in a first approximation, be inversely proportional to the ninth power of the distance between their centers, if in these centers we have positive charges $8\,e$, and around each of them eight electrons placed at the corners of a cube.

The above examples, by which it was explained how higher powers of r can be introduced in the expression for the potential energy, show also that when the potential is made to decrease so rapidly with increasing distance it will at the same time come to depend on the direction of the line joining the acting system and an external point. The factors $\frac{x'-x}{r}$ and $\frac{(x'-x)(y'-y)}{r^2}$ show this influence. In this way it is seen that by means of our "neutral atoms" we can never get a truly "central" force varying as some high power of $1/r$. The potential corresponding to such a force would have to be of some such form as β/r^5 or β/r^9, where β is a constant independent of direction.

That the electrostatic action due to a neutral atom can never have this character of a central force can be immediately shown

by describing a spherical surface around the atom and by attending to the normal components of the electric force at the points of the sphere. If the field were truly central the force E would be the same at all points of the sphere, so that the integral

$$\int E_n \, d\sigma$$

over the spherical surface would be different from zero. In reality this integral must be zero, because E also represents the dielectric displacement, the surface integral of which must be equal to the electric charge of the system within the sphere, and this is zero in the case of our neutral atom. The difficulty is that in some of his conclusions Born uses the second term of his potential energy as if it were the potential due to a really central force.

46. A Theorem in Dynamics and its Applications. We shall now pass on to the discussion of some points in Bohr's theory of spectral lines, the aim being especially to make some remarks about his "principle of correspondence," which plays an important part in the further development of his theory. By way of introduction I shall give you a theorem of pure dynamics.

Let us consider a material system under the action of conservative forces such that it can perform some periodic motion M. By this I mean simply that at two moments, between which there is a definite interval of time θ, the positions of the parts of the system are exactly the same. You have not to think merely of a simple harmonic motion; the periodic motion, at present under consideration, is more general than this.

We shall denote by E the total energy of the system, by U its potential, and by T its kinetic energy. We shall, moreover, compare the motion M with a second motion M', differing very slightly from it but belonging to those motions that can actually take place under the action of the given forces. The motion M' is supposed to be again periodic, but its period θ' may differ somewhat, though very slightly, from that of M.

We shall be led to introduce for each of the two motions the time integral

$$P = \int T\,dt$$

of the kinetic energy taken over a full period, and we shall derive a relation between the changes or variations of E and P when we pass from M to M' and the period θ; namely,

$$\frac{\delta E}{\delta P} = \frac{2}{\theta}.$$

Proof. Let x, y, z be the coördinates of one of the material points of the system, m its mass, and X, Y, Z the components of the force acting on m; then the equations of motion are

$$X = m\ddot{x} \text{ etc.}$$

Let these be the equations of the original motion M with period θ, and let us consider one such period ranging from t_0 to $t_1 = t_0 + \theta$. The period of the varied motion M' will be not θ but, say, $\theta + \delta\theta$. We shall therefore, for each moment t, which we consider in the case of the first motion, choose a varied time $t + \delta t$, where δt, the shift of our time, is a continuous function of t itself, such that when t goes through the interval θ, $t + \delta t$ ranges over the interval $\theta + \delta\theta$.

Now the position which one of the points has in the motion M' at the time $t + \delta t$ will be somewhat different from its position in the motion M at time t. We shall therefore denote the coördinates in this varied position by $x + \delta x$ etc. The infinitely small increments δx etc. will again be functions of t, and on account of the periodicity of both motions they will have the same values for t_0 and t_1.

We shall consistently use the symbol δ to indicate the difference of two corresponding quantities in the motions M and M' taken at corresponding times, and the symbol d to denote an infinitely small element of time t and all quantities belonging to this dt. This being agreed upon, you easily see the meaning of expressions like $d\delta x$ or δdx, and you see that these values are equal; also, $d\delta t = \delta dt$.

Since $dx + \delta dx$ is the distance over which in the motion M' a point travels in the direction of OX during the time $dt + \delta dt$, we have

$$(\dot{x} + \delta\dot{x})(dt + \delta dt) = dx + \delta dx,$$

or, taking into account the relation $\dot{x}dt = dx$, and neglecting products of two variations,

$$\delta dx = \delta\dot{x}dt + \dot{x}\delta dt, \quad \delta dy = \delta\dot{y}dt + \dot{y}\delta dt \text{ etc.} \tag{172}$$

We now deduce from the equations of motion that

$$\sum(X\delta x + Y\delta y + Z\delta z) = \sum m(\ddot{x}\delta x + \ddot{y}\delta y + \ddot{z}\delta z),$$

or, multiplying by dt

$$\begin{aligned} -\delta U dt &= \sum m(d\dot{x}\delta x + d\dot{y}\delta y + d\dot{z}\delta z) \\ &= d\sum m(\dot{x}\delta x + \dot{y}\delta y + \dot{z}\delta z) - \sum m[\dot{x}d\delta x + \dot{y}d\delta y + \dot{z}d\delta z] \\ &= d[\quad] - \sum m(\dot{x}\delta\dot{x} + \dot{y}\delta\dot{y} + \dot{z}\delta\dot{z})dt - \sum m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)\delta dt \end{aligned}$$

when use is made of (172).

$$\therefore -\delta U dt = d[\quad] - \delta T dt - 2\,T\delta dt. \tag{173}$$

Adding $-\delta T dt$ to both sides and changing the signs, we have

$$\delta E dt = -d[\quad] + 2\,\delta(T\,dt).$$

Now integrate this with respect to t over the period from t_0 to $t_1 = t_0 + \theta$. On the left-hand side δE is a constant, while on the right-hand side the first term gives zero and the second term gives $2\,\delta P$. Thus

$$\theta\delta E = 2\,\delta P,$$

or

$$\frac{\delta E}{\delta P} = \frac{2}{\theta}. \tag{174}$$

EXAMPLE 1. A particle subject to a quasi-elastic force under which it can perform simple harmonic motions has on the average the same amount of kinetic and potential energy,

$$P = \tfrac{1}{2}\int_0^\theta E\,dt = \tfrac{1}{2}E\theta;$$

whence, since θ is the same for all motions,

$$\delta P = \tfrac{1}{2}\,\theta\,\delta E.$$

EXAMPLE 2. A particle of mass m moves with constant velocity v between two perfectly elastic walls which it strikes normally. A new periodic motion is obtained by changing v. If l is the distance between the walls,

$$\theta = \frac{2\,l}{v},$$

$$P = \frac{2\,lT}{v} = lmv,$$

$$E = T = \tfrac{1}{2}\,mv^2,$$

$$\delta E = mv\,\delta v, \quad \delta P = lm\,\delta v,$$

$$\therefore \frac{\delta E}{\delta P} = \frac{v}{l} = \frac{2}{\theta}.$$

EXAMPLE 3. In the case of a circular motion under the action of a central force inversely proportional to r^{n+1} we have

$$U = -\frac{k}{r^n}, \quad \frac{mv^2}{r} = \frac{nk}{r^{n+1}}, \quad T = \frac{nk}{2r^n},$$

$$v = \sqrt{\frac{nk}{m}}\,r^{-\frac{1}{2}n}, \quad \theta = 2\pi\sqrt{\frac{m}{nk}}\,r^{\frac{1}{2}n+1},$$

$$P = \pi\sqrt{mnk}\,r^{1-\frac{1}{2}n}, \quad E = (\tfrac{1}{2}\,n - 1)kr^{-n},$$

$$\delta E = -\,n(\tfrac{1}{2}\,n - 1)kr^{-n-1}\delta r,$$

$$\delta P = -\,\pi(\tfrac{1}{2}\,n - 1)\sqrt{mnk}\,r^{-\frac{1}{2}n}\,\delta r,$$

$$\frac{\delta E}{\delta P} = \frac{2}{\theta}.$$

EXAMPLE 4. Finally, in the case of a rectilinear motion, the potential energy depending on one coördinate,

$$U = f(x), \quad \tfrac{1}{2}\,mv^2 + U = E \text{ (a constant)},$$

$$\left(\frac{dx}{dt}\right)^2 = \frac{2}{m}\,(E - U).$$

Assume that $U = E$ for $x = x_1$ and $x = x_2$, $(x_2 > x_1)$, then

$$\theta = 2\int_{x_1}^{x_2}\frac{dx}{\sqrt{\frac{2}{m}\,(E - U)}},$$

$$Tdt = (E - U)dt = \sqrt{\tfrac{1}{2}\,m(E - U)}dx,$$

$$P = 2\int_{x_1}^{x_2}\sqrt{\tfrac{1}{2}\,m\,(E - U)}\,dx.$$

When we pass from M to M', E changes, and so also do x_1 and x_2, but the changes of the limits x_1 and x_2 have no influence on P because the integrand is zero at the limits; therefore

$$\frac{\delta P}{\delta E} = \int_{x_1}^{x_2} \frac{dx}{\sqrt{\frac{2}{m}(E-U)}} = \frac{1}{2}\theta.$$

47. Bohr's Principle of Correspondence. Bohr supposes that only certain states of an atom, which he calls stationary states, are possible, and that emission of light only takes place when an atom changes from one of these states to another of smaller energy. There is then radiation of frequency

$$\nu = \frac{E - E_1}{h}, \tag{175}$$

where h is Planck's constant, E and E_1 denoting the initial and final values, respectively, of the energy of the atom. The mechanism of radiation is wholly obscure; its frequency, for instance, is not equal to any frequency really existing in the atom.

Let us suppose that the coördinates of the moving particles in an atom are all functions of the time with period θ. They can then be expanded in Fourier series, the frequencies of the terms being

$$\frac{1}{\theta},\quad \frac{2}{\theta},\quad \frac{3}{\theta},\quad \text{etc.}$$

If we take only the terms of a definite frequency, we have one of the harmonic components of the motion.

In the case of periodic motions the quantum condition for a stationary state is

$$P = \int_{t_0}^{t_0+\theta} T\,dt = \tfrac{1}{2}\,kh, \tag{176}$$

where k is a positive integer. The energy E is a function of k. Each transition is due to a sudden change from one value of k to a smaller value of k. Let us consider the states Q, Q_1, Q_2, etc., with energies E, E_1, E_2, etc., and values of k equal to k, $k-1$, $k-2$, etc. For the transitions $E \rightarrow E_1$, $E \rightarrow E_2$, etc. (one-jump, two-jump, etc.) we have the corresponding frequencies

$$\frac{E - E_1}{h},\quad \frac{E - E_2}{h},\quad \frac{E - E_3}{h},\quad \text{etc.,} \tag{177}$$

for which we may also write, denoting in each case by ΔE the diminution of E and by Δk that of k,

$$\frac{1}{h}\left(\frac{\Delta E}{\Delta k}\right)_1, \ \frac{1}{h}2\left(\frac{\Delta E}{\Delta k}\right)_2, \ \frac{1}{h}3\left(\frac{\Delta E}{\Delta k}\right)_3, \text{ etc.} \tag{178}$$

Now the different states of motion that can exist according to the laws of dynamics, and among which the stationary states are certain selected ones, form a continuous series, in which each of them can be characterized by the value of one variable, for which we can take E, or P, or, if we apply (176) to any state, the number k. Making this choice, we can consider the energy as a continuous function of k. Plotting E against k, we shall find a curve like that of the diagram (Fig. 35), and we can mark on it the points Q, Q_1, Q_2 etc., corresponding to the stationary states; that is, to the values k, $k-1$, $k-2$, etc.

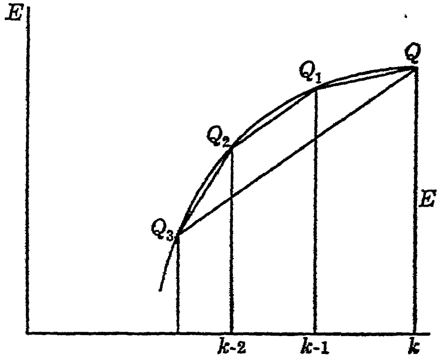

FIG. 35

The gradients $\left(\frac{\Delta E}{\Delta k}\right)_1$ etc. are the slopes of the lines QQ_1, QQ_2, QQ_3, etc. Now what are the frequencies that really occur, say in the state Q? Clearly they are

$$\frac{1}{\theta}, \ \frac{2}{\theta}, \ \frac{3}{\theta}, \text{ etc.}$$

But by virtue of (176) we have, for the transition from the state Q to another differing from it by an infinitely small amount,

$$\delta P = \tfrac{1}{2} h \,\delta k,$$

so that, by our theorem,

$$\frac{1}{\theta} = \frac{1}{2}\frac{\delta E}{\delta P} = \frac{1}{h}\frac{\delta E}{\delta k}.$$

The frequencies existing in the system are therefore

$$\frac{1}{h}\frac{\delta E}{\delta k},\quad \frac{1}{h}2\frac{\delta E}{\delta k},\quad \frac{1}{h}3\frac{\delta E}{\delta k},\quad \text{etc.}$$

Here $\frac{\delta E}{\delta k}$ is always the same and is given by the slope of the tangent to the curve at Q.

If these frequencies, which really exist in the system, are compared with the frequencies represented by (178), a close analogy between the two becomes apparent, the difference being only this, that in one case we are concerned with the slope of the tangent to the curve and in the other with the slope of the lines QQ_1, QQ_2, QQ_3 etc. In this sense, a quantum jump in which k changes by one unit ($k \rightarrow k-1$, say) corresponds to the fundamental vibration in the real motion and the jumps in which k changes by more than one correspond to the different harmonics. Now, Bohr supposes that this correspondence goes so far that there is a good deal of similitude between the light emitted in a quantum jump and the light that would be emitted if the corresponding harmonic were the only one existing and were radiating according to classical mechanics. The similitude relates to the state of polarization and to the intensity.

In this idea there is some vagueness, for in the case of every quantum jump we are concerned with two stationary states, the initial and the final one. For both of these we can fix our attention on the fundamental or primary vibration and the higher harmonics. Now we can just as well or just as little expect a correspondence between light really emitted and the harmonics in the final state as between real emission and the harmonics in the initial state. This will make no important difference, however, if the initial and final states are not too different from each other. The conclusions drawn from the principle agree with observed facts in a marvelous way.

Consider a harmonic motion of order s, the coördinates of each electron being expressed by quantities of the form

$$a \cos 2\pi s(\nu t + q),$$

where $\nu = \frac{1}{\theta}\cdot$ Now since the atom is very small compared with the wave length, it can be shown that, according to classical electrodynamics, the resulting radiation from all the electrons moving in this way would be equivalent to that from one electron having some motion of the same kind; that is, from one electron whose coördinates would be

$$a_1 \cos 2\,\pi s(\nu t + q_1), \quad a_2 \cos 2\,\pi s(\nu t + q_2), \quad a_3 \cos 2\,\pi s(\nu t + q_3).$$

This one equivalent electron would describe an ellipse, and it would always, according to classical theory, give rise to a radiation of a definite intensity and in a definite state of polarization, changing, it is true, with the direction of the radiation.

Thus, if the motion of the equivalent electron were rectilinear, we should have, for all directions of propagation, electric vibrations in a plane containing the line of motion of the electron.

When the motion of the electron is circular or elliptic, the radiation is circularly or elliptically polarized.

According to Bohr's principle the radiation produced by an s-jump will present similar states of polarization, and its intensity will be determined to a certain extent by that of the harmonic in question.

EXAMPLE 1. Suppose that a_1, a_2 and a_3 are zero, so that the s-harmonic is not present. If this is the case both for the initial and for the final motion, we can safely conclude that there will be no radiation due to an s-jump.

EXAMPLE 2. If, for instance in the case of the Stark effect, the s-harmonic is a rectilinear vibration parallel to the lines of force, we shall admit that the s-jump produces a radiation with electric forces in the plane containing the lines of force and the direction of propagation.

We must next say a few words about less simple cases, in which more than one quantum condition is required for the determination of the stationary states. Without considering cases of this kind in detail I shall make some remarks on a class that is rather general.

The quantum conditions are now of the form

$$P_1 = \tfrac{1}{2}\,k_1 h, \quad P_2 = \tfrac{1}{2}\,k_2 h, \quad \text{etc.,} \tag{179}$$

where P_1, P_2, $\cdots$ are certain quantities comparable with the one P that we had in the previous case, and where k_1, k_2, $\cdots$ are integers. The energy E of the system can be expressed as a function of these quantum numbers.

Furthermore, in the real motion we must recognize not only one fundamental frequency ν but a certain number ν_1, ν_2, $\cdots$. We mean by this that all the coördinates can be expressed as functions of the variables $\xi_1 = \nu_1 t$, $\xi_2 = \nu_2 t$, etc., these functions having with respect to all these variables the period 1. I mean that each of the coördinates is of the form $F(\xi_1, \xi_2, \cdots)$, where the function F is such that it reduces to its original value as soon as one of the variables is increased by 1, the other variables remaining constant. Now functions of this kind can be expanded in Fourier series, the structure of which is, however, less simple than it was in the former case. There are not only terms containing the factors

$$\cos 2\,\pi(\nu_1 t + \delta), \quad \cos 2\,\pi(2\,\nu_1 t + \delta'), \quad \cdots,$$
$$\cos 2\,\pi(\nu_2 t + \epsilon), \quad \cos 2\,\pi(2\,\nu_2 t + \epsilon'), \quad \cdots,$$

but also terms of the more general form

$$\cos 2\,\pi[(s_1\nu_1 + s_2\nu_2 + s_3\nu_3 + \cdots)t + \delta],$$

where s_1, s_2, s_3, $\cdots$ are integers.

It is easily seen that an expression of this kind has the properties which we assigned to $F(\xi_1, \xi_2, \cdots)$. The frequency is now

$$s_1\nu_1 + s_2\nu_2 + s_3\nu_3 + \cdots, \tag{180}$$

and this is the general expression for the frequency of one of the harmonic motions into which, in the case now under consideration, the total motion can be decomposed. We have not only harmonics of each of the fundamental modes of motion, but also motions due to the combination of two or more of them.

It should be remarked that if ν_1, ν_2, ν_3, $\cdots$ are mutually incommensurable, the total motion is not periodic. In this general case we have the relations

$$\frac{\partial E}{\partial P_1} = 2\,\nu_1, \quad \frac{\partial E}{\partial P_2} = 2\,\nu_2, \quad \cdots, \tag{181}$$

or

$$\nu_1 = \frac{1}{h}\frac{\partial E}{\partial k_1}, \quad \nu_2 = \frac{1}{h}\frac{\partial E}{\partial k_2}, \quad \cdots. \tag{182}$$

In order to indicate the nature of the correspondence which again exists between the really existing simple harmonic motions and the vibrations that are emitted by quantum jumps, it will be sufficient to take the case of only two quantum numbers k_1 and k_2. Let us consider a change in which simultaneously k_1 changes to $k_1 - s_1$ and k_2 to $k_2 - s_2$. Then, between the initial state, characterized by k_1, k_2, and the final state, characterized by $k_1 - s_1$, $k_2 - s_2$, we can intercalate the state $k_1 - s_1$, k_2. (We could also intercalate the state k_1, $k_2 - s_2$). Then, according to Bohr's rule,

$$\nu = \frac{1}{h}\left[s_1 \frac{E_{k_1, k_2} - E_{k_1 - s_1, k_2}}{\Delta k_1} + s_2 \frac{E_{k_1 - s_1, k_2} - E_{k_1 - s_1, k_2 - s_2}}{\Delta k_2}\right], \quad (183)$$

since $\Delta k_1 = s_1$ and $\Delta k_2 = s_2$. Now the fractions occurring in this expression are simply slopes, such as we had in the former case. Each one is the ratio of the change in E, produced by some change in k, to the latter change itself. We could call the two fractions k_1-slope and k_2-slope. If we represent E as a function of k_1 and k_2 by a surface, these slopes are given by the directions of lines that are easily found.

Thus we see that one of the frequencies that is really emitted is equal to $1/h$ multiplied by the sum of s_1 times the k_1-slope and s_2 times the k_2-slope. On the other hand, there is in the real motion a frequency

$$\frac{1}{h}\left[s_1 \frac{\partial E}{\partial k_1} + s_2 \frac{\partial E}{\partial k_2}\right]. \quad (184)$$

The correspondence is clearly seen.

Though the principle of correspondence has led to results that are in agreement with observed facts, even to a most wonderful degree, we must own that the relation between the motion in the atoms and the emitted radiation remains profoundly mysterious. Twenty years ago one never would have thought that we should once have a theory of the emission of light in which there are no vibrators at all. It is true that in Bohr's principle of correspondence there is question of some relation between the frequencies really existing in the atom and the frequencies emitted by quantum jumps, but we cannot fathom the deeper

meaning of this relation. Take, for instance, the conclusions to which the principle has led as to the intensity of the several spectral lines. The line produced by a three-jump, for instance, will be strong when the third harmonic in the real motion of the electrons has a great intensity. Now the intensity of the line due to a three-jump will depend on the number of times that such a jump occurs, whereas the intensity of the third harmonic has nothing to do with jumps but will be entirely known when the stationary states before and after the jump are given. The two things seem to be entirely different.

48. Remarks on the Radiation of Light. After all, we can scarcely refrain from thinking that the emission of light, though it may be brought about by a quantum jump, is due to a real vibration. The reason for this is the fact that with the light of a spectral line we can produce an interference phenomenon with a phase difference of very many, perhaps a million, wave lengths. This shows that in the rays emitted there is a regular succession of a great number of waves. This is true at a certain distance from the source of light (from the vacuum tube, for instance), and we cannot help thinking that it must be the case also in the immediate neighborhood of the radiating particle. The particle, therefore, emits light not instantaneously but during a great number of periods. This looks much like a vibration, with the period equal to that of the emitted light really existing at the place of emission.

We could, for instance, assume that in the luminescent gas there are not only the Bohr atoms but also vibrators, as Planck has imagined them, each of them having the property that it refuses to accept energy except to an amount equal to $h\nu$, where ν is its characteristic frequency. Let us further suppose that an atom can only make a quantum jump when the energy ΔE which it has to lose by it can be transferred to a vibrator. Then the jump can take place only when there is close at hand a vibrator whose frequency is $\nu = \frac{\Delta E}{h}$, so that the quantum of energy $h\nu$ required by that vibrator will be just equal to ΔE.

Thus this vibrator can take hold of the energy that the atom is to lose (making the jump possible), and it can be supposed to radiate this energy quite at leisure according to the ordinary laws of the electromagnetic field.

The assumption that the source of light contains these vibrators in sufficient number so that there are at least as many kinds of them as lines in the spectrum, is not very acceptable. It is, perhaps, more satisfactory to suppose that on the occasion of a quantum jump the atom itself is transformed into a vibrator. This would imply that the jump is by no means instantaneous, but that the atom passes from the first stationary state to the state of a vibrator, and acquires the second stationary state only at the moment at which, radiating as a vibrator, it has reached the energy of the second stationary state, and then passes into that state and ceases to radiate. This transformation of the atom into a vibrator will become, perhaps, somewhat less strange when we refer to the formula

$$\nu = \frac{1}{h}\frac{\Delta E}{\Delta k},$$

which gave us the frequency emitted by a one-jump. We can write for it

$$\nu = \frac{1}{2}\frac{\Delta E}{\Delta P},$$

where P is, as before, the time integral of the kinetic energy for a full period.

Suppose now that at a certain moment, when the atom has the energy E, and P has the value $\frac{1}{2}kh$, it is transformed in such a way, without any change in E or P, that it is able to radiate, and that, beginning from that moment, the changes of E and P are proportional to each other, the ratio of the two being such that when P has diminished to the value $\frac{1}{2}(k-1)h$, the energy has become that which corresponds to the stationary state characterized by the quantum number $k-1$. According to the theorem that was discussed in § 46 this would mean that during the period now in question the atom will move with the frequency $\frac{1}{2}\frac{\Delta E}{\Delta P}$; and since it is now supposed to radiate

according to the classical laws, the emitted light will have the same frequency, just as is required by Bohr's rule. You see that this amounts to saying that the atom can pass from the state represented by the point Q (Fig. 35) to that represented by Q_1 along the straight line QQ_1.

Of course we have the difficulty that there is also the possibility of a two-jump, from Q to Q_2. This would be the manifestation of another transformation of the atom, such that after it the point that represents the simultaneous values of $k = \frac{2P}{h}$ and E can move along the line QQ_2. The possibility, or rather the greater or less probability, of the two transformations now considered may very well be imagined to depend on the state of motion corresponding to the point Q; it could therefore be determined by the relative intensities, in that state, of the primary harmonic motion and the first harmonic. In this way it would be possible to understand in a certain degree the meaning of the principle of correspondence.

All these are very vague and perhaps idle speculations. If we do not want to venture them we can say, simply as a provisional description of the phenomena, that the quantum jump in the atom produces a state of things in which at some spherical surface surrounding the atom there are certain electromotive and magnetomotive forces acting with frequency $\frac{1}{h}\Delta E$. The demons which we imagined some time ago, and whose task it was to prevent the radiation so long as the atom remained in a stationary state, could now be charged with the radiation, an amount of energy ΔE being put at their disposal.

49. The Radiation of Angular Momentum. That, after all, the mechanism of the emission of light is not so very different from what the old theory taught us is also assumed by Rubinowicz, to whom we owe an important contribution to the theory of radiation.

One of the fundamental assumptions of Bohr's theory is that, although we do not know the mechanism of radiation, yet the

principle of the conservation of energy can be applied to it; the loss in energy of the atom is supposed to be equal to the energy that is radiated. Now in a mechanical or electromagnetic system there are other quantities that have a constant value, just as well as the energy, — that are conserved, as we may say. One of these is the moment of momentum.

According to a general theorem the moment of the forces acting on a system is equal to the rate of change of the resultant moment of momentum, these moments being taken with respect to some fixed point O. You know that the moment of a force with respect to a point O is given by the vector product $[rF]$, r being the radius vector from O to the point where the force is applied. The components of this vector product are $yF_z - zF_y$, $zF_x - xF_z$, $xF_y - yF_x$. The definition of moment of momentum is similar to that of the moment of a force.

For an electromagnetic system the theorem is as follows: If around the system we construct a sphere σ (or some other closed surface), and if we suppose certain forces F to act on the system, then the moment of these forces plus the moment of Maxwell's stresses acting at the surface (namely, forces exerted on the part within σ by the part without) equals the rate of change of angular momentum calculated for the part of the system within σ. If we denote the moment of Maxwell's stresses by $-M$, we can also say that M is the moment of momentum transferred or radiated in the outward direction across σ. If the state of things is periodic, then we have simply the result that for a full period the mean moment of the forces acting on the system is equal to the mean radiation of moment of momentum across σ.

Let us consider an atom containing a nucleus and an electron circulating around it in the plane of xy. (If we confine ourselves to this case, we shall not have to enter into the question of the quantum conditions for more complicated structures.) In Bohr's theory one of the quantum conditions is, in this case,

$$\text{Moment of momentum of the electron} = \frac{1}{2\pi} kh,$$

where k is the quantum number. If in a quantum jump k diminishes by Δk and the energy at the same time by ΔE, we have

$$\nu = \frac{1}{h}\Delta E,$$

and therefore for the diminution $\frac{1}{2\pi} h\Delta k$ of the moment of momentum we may write

$$\frac{1}{2\pi\nu}\Delta E \Delta k.$$

Whatever be the mechanism of radiation, this must be equal to the total radiation of moment of momentum. We have, therefore,

$$\frac{\text{Radiation of moment of momentum}}{\text{Radiation of energy}} = \frac{\Delta k}{2\pi\nu}. \tag{185}$$

Now if we know the way in which the radiation takes place, we can calculate the ratio on the left-hand side.

Rubinowicz performs the calculation * on the supposition that the radiation goes on in much the classical way. The computation is also to be found in Sommerfeld's book.† If we calculate the radiation of moment of momentum by determining the above quantity M, for instance, for a large sphere surrounding the particle, we are led to lengthy calculations. In fact, since in the expressions for M the coördinates x, y, z of a point of the sphere occur, we must use values of Maxwell's stresses ‡ which are accurate up to terms of order $1/r^3$. Therefore, since they are quadratic functions of $E_x, \cdots H_z$, it is not sufficient to consider in these components only the parts proportional to $1/r$, as we can do in calculating the flow of energy, but we must take into account terms proportional to $1/r^2$.

We may, however, follow another course. Whereas Sommerfeld and Rubinowicz consider radiation to be such as can be

* *Phys. Zeitschr.*, Bd. 19 (1918), pp. 441, 465.

† Atombau und Spektrallinien, 2d ed., p. 387.

‡ The radiation of angular momentum was first calculated from the moment of Poynting's vector (M. Abraham, *Phys. Zeitschr.*, Bd. 15 (1914), p. 914). See also Westphal, *Jahrbuch d. Radio-aktivität u. Elektronik*, Bd. 18 (1921–1922), p. 81. This method gives the same result as the other, but it depends on the assumption that angular momentum flows with the velocity of light. This is true for the Hertzian dipole when the terms of orders $1/r^2$ and $1/r^3$ in the momentum are retained. — ED.

expressed by certain values of the scalar and vector potential, we can just as well suppose it to be due to some cause acting at the center of emission, to periodically changing electromotive forces, for instance, or to a periodic motion of an electron radiating according to the old theory. These different suppositions are equivalent to each other.

We shall use the latter of them. Let an electron perform an elliptic vibration in the plane xy, the axes of the orbit being along OX and OY, and the motion being given by

$$x = a\cos(nt + q), \quad y = a\sin(nt + q). \tag{186}$$

We shall maintain this motion by external forces X and Y. These must be equal and opposite to the radiation resistance. Hence

$$\left.\begin{aligned} X &= -\frac{e^2}{6\,\pi c^3}\dddot{x} = -\frac{ae^2n^3}{6\,\pi c^3}\sin(nt+q) \\ Y &= \frac{be^2n^3}{6\,\pi c^3}\cos(nt+q) \end{aligned}\right\}. \tag{187}$$

The radiation of energy per unit of time is the sum of the flows that are due to the x-vibrations and the y-vibrations. These flows are known from a consideration of the forces. We can also say that they represent the work of the force (X, Y) per unit of time.

Let us replace $\frac{e^2n^3}{6\,\pi c^3}$ by s, so that

$$X = -\,as\sin(nt+q), \quad Y = bs\cos(nt+q).$$

The radiation of energy (= work of the force per unit of time) is

$$\overline{X\dot{x}} + \overline{Y\dot{y}} = \tfrac{1}{2}(a^2+b^2)sn.$$

Furthermore, instead of the flow of the moment of momentum, we can take the moment of the force (X, Y). This is

$$Yx - Xy = abs.$$

Substituting these values in (185) and introducing $2\,\pi\nu = n$, we find

$$\Delta k = \frac{2\,ab}{a^2+b^2}. \tag{188}$$

Now $2\,ab$ can never be greater than $a^2 + b^2$; it can at the best be equal to it. Therefore Δk can never be greater than 1. This shows that the principle of the conservation of moment of momentum can also lead, just like Bohr's principle of correspondence, to the exclusion of some quantum jumps. Successful applications of this have been made.

We may add the remark that in the above calculation it was not necessary to speak of the force (directed toward O and proportional to the distance) which must act on the electron in order to make it perform the elliptic motion, both the work of the force during a full period and its moment with respect to O vanishing.

50. Interference and the Quantum Theory. I tried to explain to you how the production of light by quantum jumps can perhaps be reconciled with our old views concerning radiation, so that these would hold as to the constitution of the emitted radiation. But the question arises, Can this constitution be really just what we have thought; that is, can there be a propagation according to Maxwell's laws, with a tendency to spread out in all directions and the impossibility of a lasting concentration of energy?

You know that phenomena like those of photo-electricity have led Einstein to his hypothesis of light-quanta. According to this, quantities of energy equal to $h\nu$ would be concentrated in small spaces, moving with the speed of light; they would even *be* light and would produce all optical effects. In this way we can understand that even very feeble light can give to an electron the amount of energy $h\nu$, for the smallness of the intensity would be due to the small number of quanta which it contains, the magnitude of each remaining the same. So we should escape the difficulty which, in the case of wave-motion, arises from the continual spreading out and weakening of the energy.

The hypothesis of light-quanta, however, is in contradiction with the phenomena of interference. Can the two views be reconciled? I should like to put forward some considerations about this question, but I must first say that Einstein is to be given credit for whatever in them may be sound. As I know

his ideas concerning the points to be discussed only by verbal communication, however, and even by hearsay, I have to take the responsibility for all that remains unsatisfactory.

Let us suppose that in the emission and propagation of light there is something that conforms wholly to Maxwell's equations, but that it has practically no energy at all, the electric and magnetic forces being infinitely small. Then in this, let us say, Fresnel radiation we shall have the ordinary laws of reflection, interference, and refraction, but we shall see nothing of it. On a screen you will have something like an undeveloped photographic image.

We can now imagine that in the production of light this Fresnel radiation is accompanied by the emission of certain quanta of energy that are of a different nature. Although their precise nature is unknown, we may suppose that energy is concentrated in small spaces and remains so. These quanta move in such a way in our "pattern" that they can never come to a place where in this pattern there is darkness. In thus traveling from the source outward each quantum has a choice between many paths. The probability of following different paths is proportional to the intensity of the radiation along these paths in Fresnel's radiation.

Now in all real cases the act of emission is repeated a great many times. Suppose it is repeated N times, and let the Fresnel radiation be the same in these different cases. Then we shall have N quanta moving in this pattern, and if their number is very great and the probability of following different paths as stated, the number of quanta coming on different parts of a screen on which we observe an interference phenomenon will be proportional to the intensity which we have in Fresnel's pattern.

These considerations can easily be extended. Take, for instance, polarization. The polarization will be in the Fresnel pattern, not in the quanta, but the quanta will illuminate a screen or a photographic plate or our retina to exactly the degree determined by the classical theory.

When light falls on the surface of a piece of glass, there is a partition between the reflected and refracted parts. The

probability of the quantum's following one path or another is determined by the well-known formulæ of Fresnel for the intensities of the reflected and the refracted light.

Suppose that in an elementary act of radiation there are a million waves; these exist in Fresnel's pattern; but the quantum of energy can have any place in the train of waves, either near the front or near the rear of these waves.

If we have an ordinary beam of light consisting of the superposition of a great number of elementary beams, we have quanta in great number distributed all through the space occupied by the beam.

51. Motion of the Light-Quanta in a Beam of Light. These considerations can be developed to a certain extent.

1. We shall assume, in the first place, that in light of the frequency ν the energy of a quantum is $h\nu$. Now we found in § 39 that between the energy ϵ of a system moving with the velocity v and its momentum there is always the relation

$$g = \frac{\epsilon}{c^2} v \qquad (g \text{ and } v \text{ vectors})$$

and we shall apply this to the quanta. Thus, if the quantum moves with the velocity v, it will have a momentum

$$g = \frac{h\nu}{c^2} v$$

in the direction of its motion. If the velocity is equal to that of light, the momentum becomes

$$\frac{h\nu}{c}. \tag{189}$$

2. Consider a beam of light represented by

$$E_x = \alpha \cos 2\,\pi\nu\left(t - \frac{z}{c}\right), \quad H_y = \alpha \cos 2\,\pi\nu\left(t - \frac{z}{c}\right).$$

The mean value, for a sufficient length of time, of the energy per unit of volume is $\frac{1}{2}\,\alpha^2$. The number of quanta per unit of volume must therefore be

$$N = \frac{\alpha^2}{2\,h\nu}. \tag{190}$$

Moving with the velocity c, they will carry across a normal section the amount of energy

$$S = Nch\nu = \tfrac{1}{2}\,\alpha^2 c \tag{191}$$

per unit of area and unit of time. This accounts for Poynting's flow. The momentum per unit of volume is $\alpha^2/2\,c$, and this may also be ascribed to the quanta, each of them having the momentum given by (189). It should be noticed that if the light were strong we could suppose the number of quanta in unit of volume to be exactly

$$\frac{1}{2\,h\nu}(E^2 + H^2),$$

and this would enable us to account for the flow of energy and the momentum in all their rapid variations. But in the case of a feeble beam, in which the quanta are scarce, there can be question only of a mean number, to be defined as follows. Let dV be an element of volume and let it be traversed, during a long time T, by a number of quanta that are within the element during the intervals $\tau_1, \tau_2, \tau_3, \cdots$. Then

$$N = \frac{\sum \tau}{T\,dV}.$$

In what follows we shall always speak of mean values taken in this sense, and of similar mean values of the energy, the flow of energy, etc.

3. Next, using the formulæ given in § 31 and § 34, consider the effect of a relativity transformation from x, y, z, t to x', y', z', t'. The beam of light becomes

$$E'_x = \alpha' \cos 2\,\pi\nu'\left(t' - \frac{z'}{c}\right), \quad H'_y = \alpha' \cos 2\,\pi\nu'\left(t' - \frac{z'}{c}\right),$$
$$\nu' = (a - b)\nu, \quad \alpha' = (a - b)\alpha.$$

For the quantum (cf. § 36)

$$g'_z = a g_z - \frac{b}{c}\epsilon, \quad \epsilon' = a\epsilon - bcg_z,$$
$$g'_z = \frac{a}{c}h\nu - \frac{b}{c}h\nu = \frac{1}{c}h\nu',$$
$$\epsilon' = ah\nu - bh\nu = h\nu'.$$

In a crystal, quanta must move in the direction of the rays and with the ray-velocity. In dispersive media the quanta must move with the group-velocity, for if the beam of light is limited by two planes, so that the train of waves has a front and a rear, quanta must remain in the limited beam to which they belong. This is a statement that is preserved in form by a relativity transformation, for it can be shown that the group-velocity is transformed exactly in the same way as the velocity of a particle.

Consider an isotropic dispersive medium at rest. The quanta move with group-velocity w. Their energy is $h\nu$ and their momentum

$$\frac{w}{c^2}\,\epsilon.$$

In a moving ponderable body the relations are somewhat more complicated.

52. Motion of the Light-Quanta in Superposed or Intersecting Beams of Light. In the case of a single beam of parallel rays there can be no uncertainty as to the motion that is to be ascribed to the quanta. Now one might think at first sight that when two (or more) beams are superposed, the quanta of each beam will move independently of the other. Although, as we shall presently see, this assumption cannot be maintained, it is interesting to examine some consequences that may be drawn from it.

Let us first consider the reflection of light by a plane mirror. For simplicity we take the case of normal incidence and suppose the mirror to be perfect. If the mirror were at rest everything would be very simple. The amplitude of the reflected rays would be equal to the amplitude of the incident rays. Each quantum is reflected with the same energy but with its momentum reversed. This produces exactly the pressure on the mirror that can be calculated in the well-known way from the electromagnetic momentum of the incident and the reflected beam.

Now black radiation is very much like a gas. In a cavity filled with it quanta are moving in all directions with the

velocity c, and thereby produce a pressure that can be calculated in much the same way as the pressure of a gas. We have only to consider a plane in the cavity and to determine the total momentum, taken in the direction of the normal, which the quanta carry across this plane per unit of time and unit of area. If N is the number of quanta per unit of volume, the number of those that pass through the plane in directions making with the normal an angle between θ and $\theta + d\theta$ will be

$$\tfrac{1}{2}\, Nc \sin\theta \cos\theta\, d\theta,$$

the sign of this expression indicating whether the motion is toward the positive or toward the negative side. Since the component in the direction of the normal of the momentum of a quantum is

$$\frac{h\nu}{c}\cos\theta,$$

positive or negative as the case may be, we easily find for the pressure

$$\tfrac{1}{2}\, Nh\nu \int_0^{\pi} \sin\theta \cos^2\theta\, d\theta = \tfrac{1}{3}\, Nh\nu, \tag{192}$$

equal, as it ought to be, to one third of the energy per unit of volume.

In connection with black radiation there is an interesting problem. From a consideration of the entropy Einstein has concluded that there must be small fluctuations in the density of the energy. These fluctuations are found to be of two kinds: in the first place, we have the irregularities in the distribution that can be explained by the classical theory of electromagnetic waves; and in the second place, superposed upon these irregularities, certain fluctuations that can be explained only by some form of quantum theory. The deviations of the first kind will be found to exist already in the Fresnel pattern, and those of the second kind may be expected on the assumption that for a given Fresnel pattern there are small departures from the expression (190) which we found for the number of quanta per unit of volume.*

Let us finally consider normal reflection by a receding mirror. This case can be derived from a reflection by a stationary

* See Note 14, Appendix.

mirror by the relativity transformation. The result is that if v is the velocity of the mirror, the frequency is altered to $\frac{c-v}{c+v}\nu$. The energy of each quantum is altered in the same ratio, but their number remains the same. It is needless to say that the diminution of the energy of the quanta corresponds to the work that is done by the light-pressure on the receding mirror. Similarly, the energy of the quanta can be increased by means of a mirror that is moving toward the incident rays. A production of a quantum takes place only when light is emitted, and quanta disappear only when they are absorbed, producing the effects for which they have been invented.*

53. New Assumption concerning the Motion of Light-Quanta. The hypothesis which we adopted in the last paragraph can never lead us to an explanation of the phenomena of interference. If, for instance, the waves emitted by a luminous point fall on two narrow openings in an opaque screen, we should have behind that screen streams of quanta traveling from the openings onward in all directions. A point at some distance would receive quanta from both openings, and we cannot understand how there can be a dark interference band, for we can hardly imagine two quanta to destroy one another, and even if we could, there would be the difficulty that when the total number of quanta is small, the point in question will not be reached at the same instant by two quanta, one from each opening.

We must therefore try to form an image of the motion of the quanta in such a way that it is determined by the Fresnel pattern, such as we find it to be after having taken into account the coexistence of the different beams. Now this can be done to a certain extent. Confining ourselves to light of a definite frequency and to phenomena that are stationary for a considerable lapse of time, we can calculate according to the classical theory, for any point in the field, the mean values

$$\tfrac{1}{2}(\overline{E^2}+\overline{H^2}) \text{ and } \overline{S}$$

* See Note 15, Appendix.

of the resultant electromagnetic energy and of Poynting's flow, $\overline{S}$ being a vector whose components are the mean values of S_x, S_y, S_z. This being done, we have to suppose that the number of quanta per unit of volume is

$$N = \frac{1}{2\,h\nu}(\overline{E^2} + \overline{H^2}), \tag{193}$$

and that their mean velocity is given by

$$\bar{v} = \frac{\overline{S}}{h\nu N}. \tag{194}$$

The words *mean velocity* now refer to the averaging of the velocities v_1, v_2, v_3, $\cdots$ of the quanta that are found successively in a definite element of volume dV, or, more accurately, $\bar{v}$ is defined by the vector equation (of § 51)

$$\bar{v} = \frac{\sum v\tau}{T\,dV}. \tag{195}$$

It is clear that, on these assumptions, a dark interference band, at which we have $\overline{E^2} + \overline{H^2} = 0$ and $\overline{S} = 0$, will not be reached by the quanta. Moreover, it is easily seen that the distribution of the quanta determined by (193) will not be disturbed by their motion. Indeed, we have

$$\frac{\partial \overline{S}_x}{\partial x} + \frac{\partial \overline{S}_y}{\partial y} + \frac{\partial \overline{S}_z}{\partial z} = 0,$$

and therefore
$$\frac{\partial (N\bar{v}_x)}{\partial x} + \frac{\partial (N\bar{v}_y)}{\partial y} + \frac{\partial (N\bar{v}_z)}{\partial z} = 0. \tag{196}$$

The latter equation shows that the numbers of the quanta entering and leaving an element dS will be equal.

Of course the motion of the quanta is not fully determined by our assumption concerning their mean velocity. The simplest way to interpret equation (194) would be to suppose that of the total number given by (193)

$$\frac{1}{4\,h\nu}\left(\overline{E^2} + \overline{H^2} + \frac{2}{c}\,\overline{S}\right) \tag{197}$$

quanta move in the direction of $\overline{S}$ with the velocity c, and

$$\frac{1}{4\,h\nu}\left(\overline{E^2} + \overline{H^2} - \frac{2}{c}\,\overline{S}\right) \tag{198}$$

with the same velocity in the opposite direction. By this we are led back in simple cases to what has been said in §§ 51 and 52.

In a single beam we have $\overline{E^2} = \overline{H^2}$ and $\overline{S} = c\overline{E^2}$, so that all the quanta move in the direction of propagation. On the other hand, when light is normally reflected by a perfect mirror, $\overline{S} = 0$. One half of the quanta then go in one direction and the other half in the opposite direction.*

I may add that since $S = c[E \cdot H]$ the expression (198) can never be negative.

Considerations that are somewhat like the above have recently been put forward by Emden.† He even goes so far as to make the quanta the bearers of the frequency, endowing them for this purpose with a rotation. As to the interference, Emden thinks, or rather hopes, that it may be possible to explain it by statistical considerations.

Before leaving the subject I must call your attention to some outstanding more or less serious difficulties.

1. According to what has been explained the quanta can account for the distribution of energy, for Poynting's flow, and therefore also for the momentum of light motion. But what about Maxwell's stresses?

We may apply to the quanta what was said in § 37 of a system of molecules and we can therefore say that by their motion they produce a system of stresses X_x, X_y, etc. equal to mean values of $v_x g_x$, $v_y g_x$, etc. multiplied by $-N$. In the case of a single beam of light we are thus led to the mean values of

* I have developed these considerations somewhat farther than I did in my lectures. I had at first said that the motion of the quanta could conceivably be always in the direction of Poynting's flow, but one of the Pasadena physicists, Mr. Sinclair Smith, immediately made the remark that when light is reflected by a perfect stationary mirror there are standing waves in which at some places there is no flow of energy at all, so that the quanta would not know how to move, or perhaps would not be allowed to move at all. (Thus, if the mirror is $x = 0$ and the field is given by

$$E_y = \cos pt \sin \frac{px}{c}, \quad H_z = -\sin pt \cos \frac{px}{c}, \quad E_x = E_z = H_x = H_y = 0;$$

we have
$$S_x = -\frac{c}{4} \sin 2pt \sin \frac{2px}{c}, \quad S_y = 0, \; S_z = 0,$$

so that $S_x = 0$ when $2px = cn\pi$, n being an integer. — Ed.)

† *Phys. Zeitschr.*, Bd. 22 (1921), p. 513.

Maxwell's stresses. But I fear that this is not true generally. Closely connected with this is the question of how far the way in which quanta are guided by the Fresnel pattern can be further elucidated.*

2. When a relativity transformation is applied to two beams of light intersecting at a certain angle, one is led to a rather complicated state of things. The frequencies become different and in the space common to the beams the groups of quanta (197) and (198) which initially moved with the velocity c in opposite directions come to travel (still with the velocity c) in directions that are not exactly opposite each other. Moreover, the magnitudes of the two groups of quanta become unequal and will not necessarily correspond to the two frequencies. So there is found to exist a certain connection between the quanta of beams of different frequencies. It is to be understood that these complications are not found in those parts of space where the beams are outside each other.

3. One can see no reason why the theory of quanta should not be applied to electromagnetic vibrations slower than those of light. Now consider the space between the plates of a condenser. By a continuous transition we can pass from alternating charges to the limiting case of a constant charge, but it is difficult to say what becomes of the quanta. In the electrostatic field their number ought to become infinite and the energy of each of them infinitely small. The tension along the lines of force between the plates, which already exists when the charge is still alternating, can never be caused by moving quanta, any more than it could by moving molecules.†

4. In the phenomenon of anomalous dispersion the course of the rays of light and the distribution of light and darkness are dependent on the absorption which we must presume to take place in quanta. Thus the Fresnel pattern, which has to guide the quanta, would be determined by something in which the quanta are already involved.

* See Note 16, Appendix.

† If in the electromagnetic field $H = 0$, $E_y = E_z = 0$, we have $X_x = \frac{1}{2} E_x^2$, but the stress component X_x derived from moving quanta or moving molecules is negative.

54. Heat Motion in Solid Bodies. We shall now consider the internal motions existing in a solid body at a given temperature and constituting what we call its heat motion. The method followed by Debye is that of standing waves; they represent the different degrees of freedom of the system, and to each of them we allot an amount of energy that is a known function of temperature, involving Planck's constant h. This procedure leads to an interesting mathematical problem and also to a difficulty. It is clear that the heat motion must be the same throughout the solid body, and the total energy must be proportional to the volume of the body and independent of its shape. We must of course find the same value for the specific heat, whether the body is a parallelepiped, a sphere, or an ellipsoid. Weyl * has really proved the theorem, at any rate for the case of wave-lengths that are very small compared with the dimensions of the body, and this is the case with which we are concerned.

The difficulty to which I alluded is this: The different standing waves that represent the degrees of freedom extend, each of them, all over the body. Thus it seems that when, for instance, a gaseous molecule strikes the body at a point P, and when the body has to decide to what extent it will accept the energy that is offered to it, not only the parts of the body near P, but also the more distant parts, have something to say in the matter. This is not very natural.

If, instead of standing waves, we consider progressive waves, the feature of the state of radiation to which I first called your attention, namely, uniform distribution, is found at once, and the difficulty which I pointed out disappears.

In order to make this clear I shall consider a crystalline body whose particles are distributed in a regular way, being at the points of a space lattice, and are also regularly orientated. Now if between the constituent particles there are certain forces which, after a displacement, tend to drive the particles back toward their positions of equilibrium, we can have states

* *Math. Ann.*, Bd. 71 (1912), p. 441. See also M. v. Laue, *Ann. d. Phys.*, Bd. 44 (1914), p. 1197.

of motion consisting of the propagation of plane waves, the wave-length varying from an infinitely great value to the smallest ones that are possible and that are of the order of magnitude of the distance between the molecules. At least this is true in the case of the elastic waves, in the propagation of which the above-mentioned forces play a part.

It must be remarked that if the particles of the body carry electric charges, the elastic waves are at the same time electromagnetic waves; we shall include in our reasoning all electromagnetic waves that can exist in the body, even far into the ultra-violet. When we consider the waves whose lengths are comparable to the molecular distances, we shall not replace the body by a homogeneous distribution of matter, but shall leave it as it is. Furthermore, it is to be understood that in what follows there is no question of all possible waves that can exist in the body, but only of those that are independent of each other.

If, for instance, we are concerned with harmonically vibrating waves, there will be for a chosen frequency and direction of propagation no more than three directions of vibration that are mutually independent.

Suppose that we are given the direction of the normal to the waves and the frequency ν. Then the velocity of propagation of the waves will have a definite value v, depending on direction. It can also depend on the frequency if there is dispersion, and this will be the case in a high degree when we come to wave-lengths comparable to molecular distances. If there is a dispersion, we can introduce the group-velocity of the waves, which we shall denote by w and which is related to v by the formula which we formerly deduced.

Now let the body be traversed in all directions by plane waves of different frequencies, so that we have a state of things comparable to that existing in ether filled with black radiation. Let us consider only those waves whose frequencies lie between ν and $\nu + d\nu$, and whose normals are included in an infinitely narrow cone of opening $d\omega$. Let $Q\,d\omega\,d\nu$ be the amount of energy, in so far as it belongs to the waves just specified, that

exists per unit of volume. Then it can be shown that any two bodies will be in equilibrium if Q has the value

$$Q = \frac{A}{wv^2}, \tag{199}$$

where A is a function of ν and T that is the same for all bodies.

The proof requires only some geometrical considerations and a theorem of reciprocity which is due to Helmholtz and which teaches us that if in reflection or refraction a beam of light 1 gives rise to a beam 2, whose intensity is a certain fraction of 1, the same fraction will be found when, reversing the direction of propagation, we make 2 the incident beam and ask what portion of the energy will follow the path 1. I cannot give you the proof here; I need only remark that it is not necessary for the waves to be of unlimited extent; they may be limited in the two ways that we formerly considered; but both the lateral and the longitudinal extension are supposed to be large in comparison with the wave-lengths. The surface of separation between the bodies is supposed to be plane, or at least the radius of curvature is supposed to be large compared with the maximum diameter of a cross-section of each beam.

We know that for the black radiation in the ether

$$A = \frac{h\nu^3}{e^{\frac{h\nu}{kT}} - 1}. \tag{200}$$

Therefore we shall have thermal equilibrium if in any solid body

$$\frac{h\nu^3 d\omega\, d\nu}{wv^2\left(e^{\frac{h\nu}{kT}} - 1\right)} \tag{201}$$

is the energy per unit of volume belonging to the interval $d\omega\, d\nu$. This formula includes the well-known result that in a ponderable diathermous body the density of the radiant energy is proportional to the cube of the index of refraction (namely, if $w = v$). It can be applied to the theory of the specific heats, and we should simply have to integrate over all directions and all frequencies.

We can go back from it to Debye's theory if we consider the body as perfectly homogeneous and non dispersive, so that $w = v$. Then, integrating (201) with respect to $d\omega$, and taking together two transverse and one longitudinal wave, we get

$$4\pi\left(\frac{2}{v_t^3}+\frac{1}{v_l^3}\right)\frac{h\nu^3 d\nu}{e^{\frac{h\nu}{kT}}-1}. \tag{202}$$

We here integrate from $\nu = 0$ to $\nu = \nu_0$, where we can choose ν_0 so that for high temperatures we get the energy that is required by equipartition.*

55. Theory of Dispersion and Absorption. Let us now return to some questions belonging to the old physics, in which formerly we thought that we were on very safe ground, but now we must confess that we are on very unsafe ground. I refer to the phenomenon of dispersion and absorption.

We have already observed that the difference between the velocity of light in the ether and in a ponderable body can be explained by supposing that electric moments are induced in the molecules. We shall consider this somewhat more closely.

For the sake of simplicity, though we shall speak of molecules and of electrons contained in them, we shall neglect some considerations that would, strictly speaking, be necessary in a case of molecular discontinuity, but that are wholly unimportant for our purpose.

Let us suppose again that light is being propagated in the direction Ox and that E_y, D_y, H_z are the only components different from zero, B being assumed to be the same as H. Maxwell's equations then give

$$-\frac{\partial H_z}{\partial x}=\frac{1}{c}\frac{\partial D_y}{\partial t},\quad \frac{\partial E_y}{\partial x}=-\frac{1}{c}\frac{\partial H_z}{\partial t},$$

from which we derive $$\frac{\partial^2 D_y}{\partial t^2}=c^2\frac{\partial^2 E_y}{\partial x^2}. \tag{203}$$

Now suppose that in each molecule there is an electric moment p_y proportional to E_y and in phase with it. This would be the

* See Note 17, Appendix.

case if in the equation of motion of an electron in the molecule we might neglect the term containing the acceleration, so that the equation would be

$$0 = -f\eta + eE_y,$$

where η is the displacement from the position of equilibrium, in the direction of Oy, and $-f\eta$ a quasi-elastic force. Then we should have, supposing each molecule to contain but one movable electron (cf. § 26),

$$\left.\begin{aligned} \eta = \frac{e}{f}E_y, \quad p_y = \frac{e^2}{f}E_y, \quad P_y = \frac{Ne^2}{f}E_y \\ D_y = \left(1 + \frac{Ne^2}{f}\right)E_y \end{aligned}\right\}, \tag{204}$$

and instead of (203)

$$\left(1 + \frac{Ne^2}{f}\right)\frac{\partial^2 E_y}{\partial t^2} = c^2\frac{\partial^2 E_y}{\partial x^2}, \tag{205}$$

our well-known equation, from which we can see that the velocity of propagation would be

$$v = \frac{c}{\sqrt{1 + \frac{Ne^2}{f}}},$$

so that, if μ is the index of refraction,

$$\mu^2 = 1 + \frac{Ne^2}{f}. \tag{206}$$

This is greater than unity because the moment of each particle has the direction of the electric force; it is also independent of the frequency of the light, and so there is no dispersion.

Now, long ago, before the days of the electromagnetic theory, dispersion was explained by different physicists (Sellmeier, Helmholtz) on the assumption that ponderable bodies contain small particles that are set vibrating by the incident light and that have a certain mass. It was natural to introduce this idea into Maxwell's theory and to assume that the particles

in question carry electric charges. By this it became conceivable that they can be set in motion by the electric forces existing in the light, which would have no grasp on uncharged particles. This was, in fact, one of the ways in which one has been led to the theory of electrons.

If m is the mass of the electron, the equation of motion becomes

$$m\frac{d^2\eta}{dt^2} = -f\eta + eE_y, \tag{207}$$

so that if

$$E_y = a\cos n\left(t - \frac{x}{v} + p\right),$$

we have, η being a simple harmonic function of t,

$$(-mn^2 + f)\eta = eE_y,$$

$$\eta = \frac{e}{f - mn^2}E_y,$$

and

$$\mu^2 = 1 + \frac{Ne^2}{f - mn^2}.$$

We can simplify this by putting

$$\frac{f}{m} = n_0^2,$$

where $n_0/2\pi$ is the frequency of the free vibrations of the electron. Then

$$\mu^2 = 1 + \frac{Ne^2}{m}\cdot\frac{1}{n_0^2 - n^2}. \tag{208}$$

Thus, if $n < n_0$, μ is greater than 1 and increases with n. This corresponds to the case of ordinary dispersion.

If $n > n_0$, we have $\mu < 1$, and this means that $v > c$. This is not in contradiction to the theory of relativity. Supposing the ratio a/b between the coefficients of the transformation to be greater than c/v, we simply find that if the velocity of waves in one system of coördinates is greater than c, the same is true in any other system of coördinates to which we may pass by means of a relativity transformation.

For the group-velocity we find

$$\frac{1}{w}=\frac{1}{v}-\frac{n}{v^2}\frac{dv}{dn},$$

$$\frac{1}{w}=\frac{1}{\mu c}\left[1+\frac{Ne^2}{m}\cdot\frac{n_0^2}{(n_0^2-n^2)^2}\right], \tag{209}$$

so that $\frac{1}{w} > \frac{1}{\mu c}$, $w < \mu c$, and, since $\mu < 1$, $w < c$.

The group-velocity is thus less than the velocity of light in ether.* In some cases a wave-velocity above the speed of light, that is, a refractive index less than unity, has really been observed, as, for instance, by Kundt in the case of silver.

The value of $v > c$ is due to the circumstance that when $n > n_0$ the electric moment has a direction opposite to that of the electric force. This may be illustrated by means of a simple experiment described in Rayleigh's "Theory of Sound."

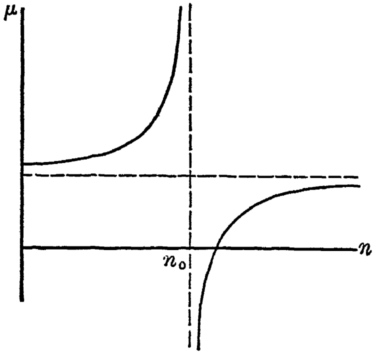

FIG. 36

For $n = n_0$ we should have an infinitely great moment and both positive and negative values of μ which are infinitely great. This is represented graphically in Fig. 36.

We cannot admit these infinite values; therefore we introduce a resistance which, for the sake of simplicity, we suppose to be proportional to the velocity of the electron. Thus the equations of motion become

$$m\frac{d^2\eta}{dt^2}=-mn_0^2\eta-g\frac{d\eta}{dt}+eE_y. \tag{210}$$

* The behavior of the group-velocity in the neighborhood of an absorption line has been discussed in some detail by A. Sommerfeld, *Ann. d. Phys.*, Bd. 44 (1914), p. 177; L. Brillouin, ibid., p. 203. See also K. F. Herzfeld, *Zeitschr. f. Phys.*, Bd. 23 (1924), p. 341. — ED.

We suppose E_y to contain a factor e^{int} which multiplies an expression independent of t, then η does so also and

$$[m(n_0^2 - n^2) + ing]\eta = eE_y. \tag{211}$$

Let us write $m(n_0^2 - n^2) = \alpha \cos \beta, \quad ng = \alpha \sin \beta,$
where α is positive; then

$$\alpha e^{i\beta}\eta = eE_y, \quad \eta = \frac{e}{\alpha} e^{-i\beta} E_y,$$

$$P_y = \frac{Ne^2}{\alpha} e^{-i\beta} E_y,$$

$$\left(1 + \frac{Ne^2}{\alpha} e^{-i\beta}\right) \frac{\partial^2 E_y}{\partial t^2} = c^2 \frac{\partial^2 E_y}{\partial x^2}. \tag{212}$$

If $$E_y = a\, e^{in(t - qx)},$$

$$q^2 = \frac{1}{c^2}\left(1 + \frac{Ne^2}{\alpha} e^{-i\beta}\right). \tag{213}$$

Suppose that N is small, then

$$q = \frac{1}{c}\left(1 + \frac{Ne^2}{2\,\alpha} e^{-i\beta}\right).$$

Put $$q = \frac{1}{v} - i\frac{k}{n};$$

then $$E_y = a\, e^{-kx + in\left(t - \frac{x}{v}\right)},$$

$$P_y = \frac{Ne^2 a}{\alpha} e^{-kx + i\left[n\left(t - \frac{x}{v}\right) - \beta\right]}.$$

The real parts of these expressions are

$$\left.\begin{aligned} E_y &= a\, e^{-kx} \cos n\left(t - \frac{x}{v}\right) \\ P_y &= \frac{Ne^2 a}{\alpha} e^{-kx} \cos\left[n\left(t - \frac{x}{v}\right) - \beta\right] \end{aligned}\right\}, \tag{214}$$

showing that k is the index of absorption. This index and the index of refraction μ are determined by the equation

$$\frac{1}{v} - i\frac{k}{n} = \frac{1}{c}\left(1 + \frac{Ne^2}{2\,\alpha} e^{-i\beta}\right).$$

Thus
$$\left.\begin{aligned} \mu &= 1 + \frac{Ne^2}{2\,\alpha} \cos \beta \\ k &= \frac{n}{c}\frac{Ne^2}{2\,\alpha} \sin \beta \end{aligned}\right\}. \tag{215}$$

The angle β determines the phase difference between the force acting on the electron and the forced vibrations.

$$\left.\begin{array}{lll} \text{For } n < n_0, & 0 < \beta < \frac{1}{2}\pi; & n << n_0, \ \beta \text{ near } 0 \\ n = n_0, & \beta = \frac{1}{2}\pi; & \\ n > n_0, & \frac{1}{2}\pi < \beta < \pi; & n >> n_0, \ \beta \text{ near } \pi \end{array}\right\}. \quad (216)$$

These results may be illustrated by the diagram of Fig. 37, in which μ (upper line) and k (lower line) are represented as functions of the frequency n. The absorption is a maximum $Ne^2/2cg$ for $n = n_0$. The index of refraction has the value 1 for this frequency and becomes a maximum and a minimum for the values of n given by

$$n^2 = n_0^2 \mp \frac{n_0 g}{m}. \quad (217)$$

For these values of the frequency the index of absorption is

$$\frac{Ne^2}{2cg} \cdot \frac{n^2}{n_0^2 + n^2}, \quad (218)$$

or, in the case of a narrow absorption band, about half the maximum value.

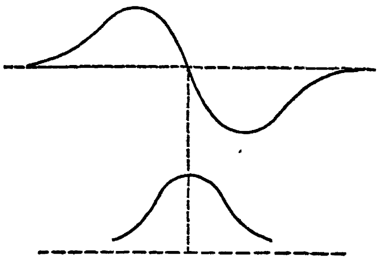

FIG. 37

The gas which we have now considered may be mixed with other gases, and in that case each of the gases will contribute to the refractive power. Each component tends to increase the index of refraction for values of n smaller than its n_0, and to lower it for values of n higher than its n_0. If we suppose that the majority of the gases in the mixture have their n_0's in the ultra-violet, then in the visible spectrum there will be a $\mu > 1$, and this will rise toward the violet. But if there is a component having its n_0 somewhere in the visible spectrum, it will raise the value of μ for frequencies a little below its n_0, and lower it for frequencies a little above its n_0.

How far these changes will extend depends on the width of the absorption band. In some cases great differences in refractivity can arise in this way. The dispersion is "anomalous" because it can happen that for a certain value of n the index of refraction is greater than for a somewhat higher value of n;

but we shall use the term *anomalous dispersion* simply to denote the dispersion corresponding to values of n in the neighborhood of n_0.

We can summarize the theory by saying that the dispersion of a substance is related to the existence of the particular modes of vibration of that substance. These show themselves in the absorption bands produced by the substance. But we must not think of absorption as a necessary condition of dispersion. The periods n_0 of the free vibrations are necessary, but the existence of these vibrations would call forth the changes which we considered even when there were no resistances. As a matter of fact, the dispersion of the air in the visible spectrum must be considered as due to the existence of an absorption band in the ultra-violet, but the absorption for yellow light, for instance, is very feeble. We must introduce a factor of resistance g when we want to consider the propagation of light of frequency n_0 or nearly n_0, but we can neglect that factor, for instance, in the case of yellow light in air.

Our formulæ show that at a distance from n_0 where absorption is imperceptible the influence on μ may still exist.

The anomalous dispersion in the neighborhood of $n = n_0$ has been observed by many physicists, and the general conclusions drawn from the theory have been verified. Therefore even though, as we shall soon see, the foundation of the theory is nowadays much shaken, we may perhaps still draw conclusions from it.

This has been done in a very ingenious way by Professor Julius,* who has made interesting applications to solar physics. The question as to how far anomalous dispersion really has the importance which Julius ascribes to it is, however, not wholly decided, and many astronomers do not share his views. The effect of the anomalous dispersion will, after all, depend on the way in which the different substances are distributed on the sun, and especially on the irregular gradients in density which exist in the sun's atmosphere. These are

* W. H. Julius, *Proc. Acad. Amsterdam*, Vol. 12 (1909), pp. 266, 446; *Astrophys. Journ.*, Vol. 31 (1910), p. 419.

things of which we know hardly anything, and so it is clear that it is difficult to predict phenomena with any great certainty.

I shall not dwell at length on the questions that now present themselves, as I am not sufficiently competent to pronounce a judgment, but I should like to explain to you the fundamental idea of Julius, about which there can, in my opinion, be scarcely any doubt. For the sake of clearness I shall simplify it as much as possible.

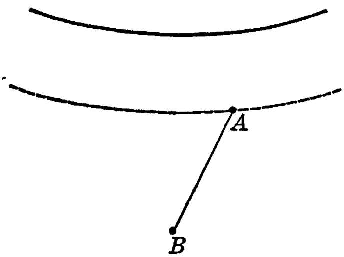

FIG. 38

Let us suppose the photosphere to be a black body at a high temperature, and let it be surrounded by an atmosphere extending to a certain height. The question is, What will be the intensity of some particular frequency n that is propagated toward some observer along the line AB (Fig. 38), for instance? In considering it we shall confine ourselves to the radiation of the photosphere, and to the refraction which it can undergo in the sun's atmosphere.

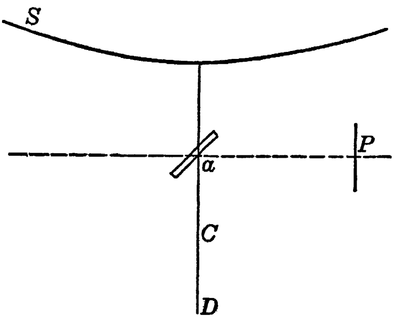

FIG. 39

Let me recall to you, to begin with, that if there were no atmosphere at all, and if the photosphere were radiating as a black body, we should see it as a disk of uniform surface brightness I_0. If before such a sphere S (Fig. 39) we were to place a plate of glass a, for instance, we should see it on the disk as a somewhat obscure patch. Indeed, some of the rays would be reflected by the plate and would not reach our eyes. If now, however, at P there were a luminous disk of the same brightness as S itself, say a body composed of the same substance as S and kept at the same temperature, the loss of light of which we spoke just now would

be made up for by light issuing from P and reflected by the plate of glass in the direction CD toward the observer. The intensity of the part of the image that is covered by the plate will be equal to that of the surrounding part, so that the plate cannot be discerned. Simple observations in everyday life will confirm this.

We can put the question most clearly, perhaps, if instead of a propagation in the direction CD we imagine one in the inverse direction DC. Suppose that a ray of light goes toward S along this line. This ray will be partially reflected and partially transmitted by our plate. Now if all the light that arrives along DC falls, after all, on a surface of the brightness of S, as it will do if P is present, and so is absorbed by the supposed black body, then, conversely, in the case of propagation toward the eye, we shall have the definite intensity which I call I_0.

If, however, part of the light arriving along DC does not fall on a radiating surface, as will be the case if P is removed, we shall miss exactly, in the radiation toward the eye, the part that was just contributed by P, and we shall have a smaller intensity.

Now consider the case to which Fig. 38 refers. Let there be on the sun's atmosphere an irregular distribution of differently refracting substances. A ray of light coming along BA can be deviated more or less from a straight line, and it may be that the beam BA is divided into several parts. Now if all these paths lead to the photosphere, but to different places in it, P_1, P_2, P_3, $\cdots$, each of these places will, by its radiation, contribute toward the total radiation that is propagated in the direction of AB, and so the intensity will again be I_0, just as if there were no atmosphere at all.

It will, however, be less when some of the paths along which the beam BA goes onward do not reach the photosphere but are directed away from the sun. Then, in the light which we perceive along AB, just that light will be wanting that could have been produced by some auxiliary luminous body such as P in Fig. 39.

Therefore the question What will be the intensity observed in the direction AB? is reduced to the question To what extent are rays of light directed toward the sun prevented by the

refraction in its atmosphere from reaching the sun's surface and deflected away from it? If this can occur to a perceptible extent, it will take place in a higher degree when BA is directed toward the sun's limb than when it points to the center. Indeed, in the second case a deflection through a greater angle would be required than in the first.

The inhomogeneity of the sun's atmosphere that may cause a refraction may consist in a change of composition from one place to another or simply in a change of density. Let us, for the sake of simplicity, consider the second case. Then we shall have a mixture of definite composition producing a general refraction depending on the density. But if n_0 is the frequency of the free vibrations of one of the substances, the refractive index for frequencies somewhat below n_0 will be greater than when that frequency did not exist, and for frequencies a little above n_0 it will be somewhat less. Now if the refractivity is altered, the differences in refractivity caused by a change of density will be altered in the same direction, and so we see that the causes which diminish the luminosity in the peripheral part of the sun's disk will be most efficient for frequencies somewhat lower than n_0. If the light is received in a spectroscope, we shall have a smaller intensity and even relative darkness on the red side of the frequency n_0. So Julius is led to the conclusion that the phenomenon of anomalous dispersion can change the aspect of the Fraunhofer lines, the lines being broadened somewhat toward the red, and their centers of gravity being thereby displaced in that direction.

It may be, as Julius pointed out, that on the changes of refractivity which we find for frequencies near n_0 there are superposed similar changes due to a frequency n_0' that differs little from n_0 and may belong either to the same substance as n_0 or to a different one. He has shown that under proper conditions the broadening of a line n_0 and its shift toward the red can be increased by the presence of a neighboring line n_0' on the violet side of n_0.

I must add, without going into all the details of his theory, that the refraction, and especially the great refractiveness which

occurs in anomalous dispersion, can lead not only to a diminution but also to an increase of intensity, or to the appearance of light where otherwise we should have darkness. Let AB (Fig. 40) be a line not intersecting the photosphere itself but passing at a short distance from it, so that we should receive no light along it if there were no solar atmosphere. It may be that by the irregular distribution of refractive matter in the actual atmosphere we get some radiation along BA, issuing, for instance, from the point P of the photosphere. Here again we may say that this case will occur if the rays coming from the side A are enabled by the refractions to reach the photosphere.

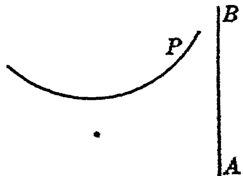

FIG. 40

According to Julius's views prominences could be produced in this way. This is certainly an interesting application of the theory of anomalous dispersion* but a decision as to the validity of the explanation must be left to those who examine solar phenomena with the utmost care, determining the position and observing the width and the aspect of the lines in the spectrum. Thus it must be left to those who continue the work of Professor Julius and to the astronomers of the Mt. Wilson and other observatories.

56. The Absorption and Dispersion of Light. I shall now make some observations about the absorption and dispersion of light as we must conceive them in the light of Bohr's theory. Let us first take absorption. It is natural to consider this as due to a process which is the inverse of radiation. The incident light, then, imparts an amount of energy $h\nu$ to the atom, thus bringing this from its original stationary state S to a new stationary state S' of greater energy. S may be the natural state which the atom assumes when left to itself. What then becomes of the absorbed energy? It may be that the atom returns from S' to S with an emission of light. Then we have resonance, much like a scattering of light, but without polarization.

* Another interesting application of the theory of anomalous dispersion has been made by Ebert in his theory of novæ, *Astr. Nachr.*, Bd. 164, p. 65. See also H. Kienle, *Phys. Zeitschr.*, Bd. 21 (1920), p. 391.—ED.

It may also be that from S' the atom returns to a state S'', different from S, with an emission of light. This phenomenon may be called fluorescence. Bohr's explanations of many phenomena observed in this domain by Wood, Dunoyer, and Rayleigh are quite remarkable. Consider, for instance, light belonging to one of the lines of the principal series of sodium. The first of these lines is the well-known line D; let the second be Q (Fig. 41). If we illuminate vapor of sodium with the light of D, we get back D as resonance radiation; but if we illuminate with Q light, we get back not only Q light but also D light.

Now both D and Q are doublets with components D_1, D_2; Q_1, Q_2. It has been found that if sodium vapor is illuminated by one of the lines D_1, D_2 separately, we get back only the light of that line. Similarly, if we illuminate by one of the Q lines, the other line of the pair will be wanting in the fluorescence but the lines D_1 and D_2 will both be present in it. All this has been very beautifully explained, but I shall not dwell upon it any longer.

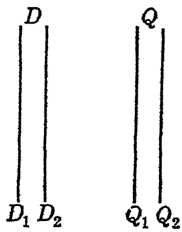

FIG. 41

As to the dispersion by itself, if we consider it apart from absorption, a theory of it can be developed to a certain extent. We have now to consider electrons not as bound by quasi-elastic forces to definite positions of equilibrium but as revolving around a nucleus. As I have already had occasion to remark, the disturbance of this motion, caused by an incident beam of light, can be calculated, at least so long as the disturbance can be considered as infinitely small. We may say that, besides the original motion, we have now a swinging to and fro called forth by the external periodic force. This is equivalent to an alternating electric moment, and it influences the velocity of propagation in the same way as such a moment would do.

Debye has worked out this idea for hydrogen,* that is, for the case of the hydrogen molecule, which he supposed to con-

* P. Debye, *Münchener Sitzungsber.* (1915), p. 1. See also A. Sommerfeld, *Ann. d. Phys.*, Bd. 53 (1917), p. 497.

sist of two hydrogen nuclei and two electrons. You probably all know the constitution which he ascribes to such a molecule (Fig. 42).

Now Debye calculated the forced vibrations that can be given to such a system, — I mean the periodic deviations from the original motion. Moreover, from his results he deduced the velocity of propagation and the index of refraction for various kinds of light. His results were in very good agreement with the observations,* so that we can say that both the refraction and the dispersion of hydrogen were satisfactorily explained. There is, however, one serious difficulty. The state of motion imagined by Debye (I mean the original circular motion) is in certain respects unstable.† This is shown by Debye's formulæ themselves. The equations which he finds for the index of refraction, or rather for $\mu^2 - 1$, contain several terms. Three of these are of the form we found above, namely,

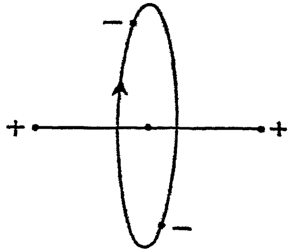

FIG. 42

$$\frac{Ne^2}{m(n_0^2 - n^2)}.$$

They have in the denominator the difference between the squares of the incident frequency and a certain quantity (here n_0) that belongs to the vibrating system and is characteristic of it. Debye therefore rightly remarks that these terms correspond to the free vibrations of the molecule.

But there are two more terms, and in these we have the sum of two positive terms in the denominator. This shows that these terms correspond to degrees of freedom for which the equilibrium is unstable. It is just as if, in the formulæ of § 55,

* M. Kirn has followed the dispersion of hydrogen up to 1854 A° and finds that the deviations of Debye's calculated values from the observed values are considerable only in the ultra-violet (*Ann. d. Phys.*, Bd. 64 (1921), p. 566). — ED.

† This has been pointed out by Miss H. J. van Leeuwen, *Proc. Acad. Amsterdam*, Vol. 18 (1916), p. 1071.

we had introduced a negative quasi-elastic force, writing for the equations of motion

$$m\ddot{\eta} = f\eta + eE,$$

which would have led us to

$$\mu^2 - 1 = -\frac{Ne^2}{mn^2 + f}.$$

Therefore it is just as if you wanted a pendulum to vibrate about a position of unstable equilibrium in which it stands vertically upward. In spite of this difficulty, which, so far as I know, has never been overcome, we may still hope that it will be possible to arrive at a satisfactory theory of dispersion by itself. But the theory of anomalous dispersion will be much more difficult. Indeed, we can scarcely see how it is possible to adapt to modern views the theory of Sellmeier and Helmholtz, which I sketched in § 55.

I may here remark that even at the time when that theory was proposed it was clear that the assumption made about the resistance $-g\dot{\eta}$ was a very crude one. When we except the radiation resistance, which was not known at the time, and which, as we can now say, would be inefficient in many cases for the production of the observed diminution of the intensity of light, no cause was known that could give rise to a resistance following the law assumed.

It was, however, possible to account for the resistance term by the hypothesis that the vibrations of the electron do not go on undisturbed for any length of time, but are destroyed at irregular intervals by a collision of one particle with another or something of the kind.* We see clearly that these actions could, just as well as the supposed resistance, prevent the amplitude from becoming infinite, and that by them the regular vibratory motion could be converted into an irregular heat motion.

In Bohr's theory the problem of anomalous dispersion, that is, of the propagation of light that is more or less absorbed by

* *Proc. Acad. Amsterdam* (1897–1898); Theory of Electrons, p. 141.

the body, is beset with many seemingly insuperable difficulties.* There is no question of vibrators at all, and if we imagine them or suppose the atom to be changed temporarily into a vibrator, we do not see how a resistance is produced. Moreover, we do not see how the vibrators are to behave toward frequencies that are somewhat greater or smaller than the frequency of their free vibrations, and it is just this (namely, the way in which vibrations are affected by frequencies somewhat below or above n_0) that was important in the old theories and is interesting from an experimental point of view.

57. The Zeeman Effect. I shall now leave the subject of dispersion in order to make a short digression on the theory of the Zeeman effect. The explanation which it finds in the new theory is based on a beautiful theorem that was proved long ago by Larmor.†

Consider a system, an atom, for instance, in which the only movable particles are electrons having all the same mass m and the same charge e. Let M be a certain state of motion which the system can perform when there is no magnetic field. Then in a magnetic field of intensity H the system can have a motion M', consisting of M and, in addition to it, a rotation around the lines of force with angular velocity

$$\omega = -\frac{e}{2\,cm}H. \tag{219}$$

In this way Larmor's theorem enables us to deduce from any motion that is possible in the absence of the field a second motion that can exist in the field. We shall call these corresponding motions.

If we want to explain the Zeeman effect by means of Bohr's theory, we have to account for the spectral lines as they are

* Attempts to overcome these difficulties have recently been made by C. G. Darwin, *Nature*, Vol. 110 (1922), p. 841; *Proc. Nat. Acad. Sci.*, Vol. 9 (1923), p. 25; A. Smekal, *Die Naturwissenschaften*, Bd. 26 (1923), Heft 43; K. F. Herzfeld, *Zeitschr. f. Phys.*, Bd. 23 (1924), p. 341.

† J. Larmor, *Phil. Mag.*, Vol. 44 (1897), p. 503; *Proc. Roy. Soc. of London*, Vol. 60 (1897), p. 514; *Cambr. Phil. Trans.*, Stokes Commemoration, Vol. 18, 1900, p. 380; Aether and Matter, p. 341.

without a magnetic field and under its influence. We have therefore in both cases to determine by quantum conditions the stationary states of motion that are supposed to be the only ones that can exist in reality, with the exclusion of all others, and we have to determine the energies of the atom in these individual states. Having found these energies, we can deduce from the differences between them, by dividing by Planck's constant h, the frequencies of the radiation produced by quantum jumps.

An explanation of the Zeeman effect is obtained if we suppose that to every stationary state in the magnetic field there corresponds (the word being taken in the above sense) a certain stationary state without the field, and also that suitable quantum conditions should be added to those which must already be satisfied when there is no field.

For the following considerations it will be sufficient to show how a spectral line L existing in the latter case can give rise to a line L' of a slightly different frequency when the field is excited. We shall have a decomposition of L if there is more than one L'.

Let us suppose that the magnetic field has the direction OZ. We have first to account for the line L. Let M_1 and M_2 be two stationary states for $H=0$, so that L corresponds to the quantum jump $M_1 \rightarrow M_2$; then the frequency of L is

$$\nu_L = \frac{E_{M_1} - E_{M_2}}{h}. \tag{220}$$

These states are wholly determined, but since there is no field the atom can have either M_1 or M_2 with all possible orientations. Suppose now that in the motion M_1 the atom has a certain moment of momentum with respect to the nucleus, and let this moment of momentum be called G_1. This is a vector attached to the atom, and the atom having the motion M_1 can be orientated in all directions, so that although the magnitude of G_1 is fixed, the component G_{1z} of G_1 along OZ can have many different values.

Now in order to explain the Zeeman effect we must, as I have already said, introduce new quantum conditions supplementing

those by which M_1 was originally determined. Let the new conditions be as follows:

When there is a magnetic field, not all the motions corresponding to the motion M_1 in its different orientations will be stationary, but only those corresponding to motions M_1 taken in such orientations that

$$G_{1z} = k_1 \frac{h}{2\pi}, \tag{221}$$

where k_1 is a positive or negative whole number, or eventually zero.

In what follows we shall understand by M_1 a motion fulfilling this condition with a definite value of k_1. Similarly, M_2 will denote a motion subjected not only to the quantum conditions already introduced but also to the supplementary condition

$$G_{2z} = k_2 \frac{h}{2\pi}, \tag{222}$$

the meaning of which will be clear.

The motions M_1' and M_2' corresponding to the particular M_1 and M_2 now specified will be stationary motions in the magnetic field, and the line L' will be due to the transition $M_1' \rightarrow M_2'$, so that we have

$$\nu_{L'} = \frac{E_{M_1'} - E_{M_2'}}{h}. \tag{223}$$

In order to compare $\nu_{L'}$ and ν_L we have therefore to compare the energies E_{M_1} and $E_{M_1'}$, and, similarly, E_{M_2} and $E_{M_2'}$. The way in which this is usually done is as follows:

Let x, y, z be the coördinates of an electron relative to the nucleus at time t in the state M_1; let $\dot{x}$, $\dot{y}$, $\dot{z}$ be the component velocities and U the potential energy due to the forces between the nucleus and electrons and between the electrons mutually. The kinetic energy will be

$$\sum \tfrac{1}{2} m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2).$$

In the state M_1', which is distinguished from M_1 only by the rotation ω, we have the same potential energy (because the

configuration is the same); but the velocity due to the rotation ω has components

$$-\omega y, \quad \omega x, \quad 0,$$

so that in the state M_1' the velocities are

$$\dot{x} - \omega y, \quad \dot{y} + \omega x, \quad \dot{z},$$

giving for the kinetic energy, when terms with ω^2 are neglected,

$$\sum \tfrac{1}{2}\, m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + \omega \sum m(x\dot{y} - y\dot{x}). \tag{224}$$

Thus
$$E_{M_1'} - E_{M_1} = \omega \sum m(x\dot{y} - y\dot{x}). \tag{225}$$

The sum on the right-hand side is, however, just the moment of momentum of the atom with respect to OZ; it is thus the quantity which we have called G_{1z}, and its value is therefore

$$k_1 \frac{h}{2\pi}.$$

We now find
$$E_{M_1'} - E_{M_1} = k_1 \frac{h\omega}{2\pi}. \tag{226}$$

In the same way we find

$$E_{M_2'} - E_{M_2} = k_2 \frac{h\omega}{2\pi}, \tag{227}$$

and from (220) and (223)

$$\nu_{L'} - \nu_L = (k_1 - k_2)\frac{\omega}{2\pi} = (k_2 - k_1)\frac{e}{4\pi cm} H. \tag{228}$$

If the difference $k_2 - k_1$ can have the values -1, 0, $+1$, we find the three values,

$$-\frac{eH}{4\pi cm}, \quad 0, \quad +\frac{eH}{4\pi cm}.$$

The three lines L' form what we call a *normal triplet*, just the triplet that was required by the old elementary theory.

Remark 1. The deduction of the equations (226) and (227) is very incomplete. I can best explain this by recalling to you the theory of a system of two conductors carrying electric currents i_1 and i_2. You know that the magnetic energy is represented by

$$\tfrac{1}{2} L_1 i_1^2 + Q i_1 i_2 + \tfrac{1}{2} L_2 i_2^2, \tag{229}$$

where L_1 and L_2 are the coefficients of self-induction, whereas Q is the coefficient of mutual induction.

The first and the last term are the magnetic energies which we should have if only the first or only the second current existed, while Qi_1i_2 is the energy due to the simultaneous existence (superposition) of the fields belonging to i_1 and i_2.

Now we have instead of i_1 our atom in the stationary state M'_1, and instead of i_2 the current which produces the field and which we can suppose to exist in a long solenoid, the axis of which is along OZ. We shall suppose this solenoid to be without resistance and to be, in fact, a perfect conductor. This assumption will not change the Zeeman effect; moreover, it has the advantage that when considering the energy we have not to trouble ourselves with the heat developed, nor with the work of the electromotive force. There must, in fact, be no electromotive force.

The energy of the state M'_1, as we have calculated it, corresponds simply to the part $\frac{1}{2} L_1 i_1^2$, of the above expression, but just because an atom has a certain rotation about the axis of the coil, it can be compared to a small coil, or winding, with its plane at right angles to OZ, or at all events not parallel to that axis. Hence there will be a second term in the energy as in (229), and if we calculate it we find that it exactly counterbalances the former expression for $E_{M'_1} - E_{M_1}$, so that at first sight it seems as if the total difference $E_{M'_1} - E_{M_1}$ would be zero, and the total difference $E_{M'_2} - E_{M_2}$ likewise zero; in that case there would be no Zeeman effect at all.

But fortunately we have still the last term in (229). When the same winding of which I spoke just now is turned, or when the electric current in it is changed, this will by induction change the intensity of the current i_2 in the great coil, and therefore we shall have a corresponding change in the last term of (229). Similarly in a magnetic field a transition from the state M'_1 to the state M'_2 with different values of G_{1z} and G_{2z} (that is, with unequal rotations of electrons about the axis OZ), will by induction cause a change in i_2 and we shall have to make a change in the last term of (229). On calculation it is found that this change is exactly

$$(k_1 - k_2)\frac{h\omega}{2\pi},$$

so that our explanation of the Zeeman effect can be maintained.

It is curious, however, that we have now been obliged to take into account an energy that is not located in the atom itself; for the energy $\frac{1}{2} L_2 i_2^2$, due to the current i_2 in the great coil, is distributed

all over the space within it. Yet this energy must be available for the emission when the atom performs its quantum jump.*

REMARK 2. Planck's constant has disappeared from our final formulæ, so it can be understood that the old theory led to the same result. In this respect there is a difference between the Zeeman effect and the Stark effect. In the latter h appears in the final formulæ; therefore it is impossible to find an explanation of the Stark effect by the old methods.

REMARK 3. It is remarkable that it was not necessary to specify the manner in which the motions M_1 and M_2 are determined by quantum conditions. Thus it seems that a theory of the Zeeman effect can be given in cases in which the spectrum has not yet been explained. Normal Zeeman triplets are observed in the case of spectral series consisting of single lines, and we can account for them without wholly understanding the origin of these lines themselves.

REMARK 4. The reasoning applied by Rubinowicz teaches us that $k_2 - k_1$ can never be greater in absolute value than 1, so that we could never expect other effects than normal triplets. More complex decompositions, in which distances between the components have been found that are in simple ratios to the distances in normal triplets, and in which distances from the middle line occur which are greater than those found in normal triplets, will at least require new assumptions such as have been tried recently by different physicists.

58. Generalized Theory of Relativity or Theory of Gravitation. In special relativity it is shown that the form of the equations by which physical phenomena are described remains the same when, by means of certain linear transformations, we pass from one system of coördinates x, y, z, t to another system. The form would be changed by an arbitrary transformation in which the new coördinates are any functions of the old ones. Now it is very remarkable that the principle which states that the form of the general equations should be the same in all systems of coördinates can be maintained if one introduces gravitational fields.

We shall begin with a short mathematical introduction. In the first place, we shall always use four coördinates, which we

* I have worked out the above considerations in an article in Marx's "Handbuch der Radiologie."

shall denote by x_1, x_2, x_3, x_4. For a reason that will appear later it would be preferable to write x^1, x^2, x^3, x^4, but there is the objection that then we should have confusion between the indices and the ordinary exponents. The coördinate x_4 will always be supposed to represent the time.

We can now imagine a four-dimensional space, say R_4, in which every system of values x_1, x_2, x_3, x_4 is represented by a point, or rather such a system *is* a point, and the totality of all possible systems of values of x_1, x_2, x_3, x_4 forms the four-dimensional space, or extension, R_4. Each event that takes place at a moment and a point is represented by a point in R_4.

If a material point, or a light-signal, is moving in the space x_1, x_2, x_3, we shall have for every value of x_4 definite values of x_1, x_2, x_3. In R_4 we shall have a line which, after Minkowski, we call the world-line of the particle or of the light-signal. The intersection of two world-lines is an encounter.

We can often explain things better by supposing a smaller number of coördinates. If there were but one space coördinate, say x_1, and the time x_4, the world-line of a moving point could be drawn in a plane by plotting x_1 against x_4, etc.

As the essence of the theory is that in changes of coördinates equations preserve their form, these changes and the relations connecting different physical quantities in one system of coördinates with the corresponding quantities in another system are of paramount interest. These relations are expressed by transformation formulæ.

Let x_1, x_2, x_3, x_4 be arbitrary functions of x'_1, x'_2, x'_3, x'_4, and conversely. The relations between infinitely small changes of x_1, x_2, x_3, x_4 and the corresponding ones of x'_1, x'_2, x'_3, x'_4 are linear and homogeneous. We shall write them in the form*

$$dx_a = \sum_{(b)} p_{ab} dx'_b. \tag{230}$$

* The sign of summation means that on the right-hand side we have four terms which can be written down if b is replaced successively by 1, 2, 3, 4. The suffix a is understood to have a definite value in all the terms of the formula. By taking $a = 1, 2, 3, 4$ we find the four equations that are contained in (230). If two or more suffixes are added to the sign Σ, this means that each of them has to be taken equal to 1, 2, 3, 4, so that if there are, for instance, three suffixes, the sum consists of sixty-four terms.

In fact, since x_a is a function of $x'_1 \cdots x'_4$, we have four equations of type

$$dx_a = \frac{\partial x_a}{\partial x'_1} dx'_1 + \cdots + \frac{\partial x_a}{\partial x'_4} dx'_4. \tag{231}$$

For the partial derivatives we can write

$$p_{a1} = \frac{\partial x_a}{\partial x'_1}, \quad p_{a2} = \frac{\partial x_a}{\partial x'_2}, \quad \text{etc.} \tag{232}$$

If, for example, there are but two coördinates, and if the relation between them is linear and homogeneous, we can write

$$x_1 = p_{11}x'_1 + p_{12}x'_2, \quad x_2 = p_{21}x'_1 + p_{22}x'_2. \tag{233}$$

We can solve these equations for x'_1, x'_2, the resulting formulæ being of the form

$$x'_1 = \pi_{11}x_1 + \pi_{21}x_2, \quad x'_2 = \pi_{12}x_1 + \pi_{22}x_2. \tag{234}$$

In general when we solve (231) for $dx'_1 \cdots dx'_4$ we find

$$dx'_a = \sum\nolimits^{(b)} \pi_{ba} dx_b. \tag{235}$$

In the above simple example the π's can readily be expressed in terms of the p's. For instance,

$$\pi_{11} = \frac{p_{22}}{p_{11}p_{22} - p_{12}p_{21}}. \tag{236}$$

Certain relations between the p's and the π's are, however, of more interest than the expressions for one set of coefficients in terms of the coefficients of the other set. These relations are obtained as follows: The values (234) must identically satisfy (233); consequently, when they are substituted in the first equation of (233), we must obtain x_1 on the right-hand side. In other words, the coefficient of x_1 must be 1; that of x_2, 0. Thus

$$p_{11}\pi_{11} + p_{12}\pi_{12} = 1, \quad p_{11}\pi_{21} + p_{12}\pi_{22} = 0. \tag{237}$$

Similarly, $p_{21}\pi_{11} + p_{22}\pi_{12} = 0, \quad p_{21}\pi_{21} + p_{22}\pi_{22} = 1.$ (238)

Again, we can also substitute the values (233) in (234) and thus obtain four more equations.

There are similar formulæ in the general case. Substituting in (230) the values (235), writing (235) in the form

$$dx'_b = \sum\nolimits_{(c)} \pi_{cb} dx_c,$$

we get

$$dx_a = \sum\nolimits_{(bc)} p_{ab} \pi_{cb} dx_c. \tag{239}$$

If on the right-hand side we want to see what is the coefficient of a definite term dx_c, we must choose c (1, 2, 3, or 4) but effect the summation with respect to the index b. We must, however, get only dx_a; therefore the coefficient must be unity when $c = a$, and zero when $c \neq a$. In other words,

$$\sum\nolimits_{(b)} p_{ab} \pi_{cb} \begin{Bmatrix} = 1 & \text{for} & c = a \\ = 0 & \text{for} & c \neq a \end{Bmatrix}. \tag{240}$$

Similarly, when we substitute (230) in (235), we get

$$\sum\nolimits_{(a)} p_{ab} \pi_{ac} \begin{Bmatrix} = 1 & \text{for} & b = c \\ = 0 & \text{for} & b \neq c \end{Bmatrix}. \tag{241}$$

These relations are frequently used.

The two variables x_1 and x_2 might be oblique or rectangular coördinates in a plane. Again, if we had two sets of three coördinates, x_1, x_2, x_3 and x'_1, x'_2, x'_3, each set being rectangular coördinates in space, the coefficients p and π would be the cosines of the angles between these axes, and (240) and (241) the well-known relations between these cosines.

We have already had occasion to observe that often there are groups of a greater or smaller number of quantities, each of which is related to one or more of the coördinates, such that together they form a kind of physical entity. For instance, in a plane, or in three-dimensional space, the two or three components of a force, or a velocity, or any other vector, form such a group. In special relativity we have already found groups of four components, — for example, the three components of momentum and the energy of a material particle. We have even found groups consisting of more components. You will remember that for a system continuously distributed over space the complex of the components of momentum, of the flow of energy, of the stress, and of the energy itself forms such

a group. Here again there are simple relations (transformation formulæ) between the quantities belonging to the group and the members of the corresponding group to which one is led when the coördinates are changed.

A group of quantities belonging together in this way is called a *tensor* when the transformation formulæ are linear and homogeneous with coefficients containing the quantities which we have denoted by p_{ab} and π_{ab}. The individual quantities forming the group are called the *components of the tensor.*

Tensors may be divided into different classes according to their transformation formulæ, those of the same class having transformation formulæ of the same form.

Let x_1, x_2 and x_1', x_2' be two sets of oblique coördinates in a plane. Let f^1, f^2 be the components of a force f in the first system, and $f^{1\prime}$, $f^{2\prime}$ those of the same force f in the second system. It is clear that the relation between these components is the same as that between the coördinates themselves (Fig. 43).

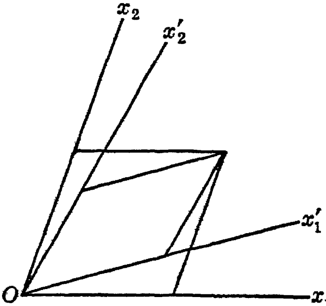

FIG. 43

The same holds for any vector. We shall therefore say that four quantities f^1, f^2, f^3, f^4 form a vector when they are connected with the corresponding quantities in the system of coördinates x_1', x_2', x_3', x_4' by relations of the same form as those connecting $dx_1 \cdots dx_4$ with $dx_1' \cdots dx_4'$. A vector is therefore characterized by the transformation formula

$$f'^a = \sum_{(b)} \pi_{ba} f^b. \tag{242}$$

The vector can be denoted by f^1, f^2, f^3, f^4 or by f^a, or simply by f. It is also called a tensor of the first rank.

To give an example of a tensor of the second rank we consider the stresses in an elastic body. The components of these are X_x, X_y, X_z, Y_x, Y_y, Y_z, etc. relative to rectangular axes. Each of these nine components is related to two axes; therefore the tensor is of the second order. In the formulæ of transformation

we have squares and products of the cosines of the angles between the two sets of axes; that is, squares and products of the p's or π's. Or again, if we take the equation of an ellipse in oblique coördinates

$$g_{11}x_1^2 + 2\,g_{12}x_1x_2 + g_{22}x_2^2 = 1, \quad (g_{21} = g_{12})$$

and make the substitution (233), we get an equation of the form

$$g'_{11}x_1'^2 + 2\,g'_{12}x'_1x'_2 + g'_{22}x_2'^2 = 1, \quad (g'_{21} = g'_{12})$$

where $\quad g'_{11} = p_{11}^2g_{11} + p_{11}p_{21}g_{12} + p_{21}p_{11}g_{12} + p_{21}^2g_{22}$ etc.

The four quantities g_{11}, g_{21}, g_{12}, g_{22} form a tensor of the second rank, and the last equation is one of the transformation formulæ.

This leads to a very important question. Suppose that we have a quantity of the second degree in the four increments dx_a, namely,

$$Q = \sum\nolimits^{(ab)} g_{ab}dx_adx_b, \tag{243}$$

the coefficients g_{ab} being functions of x_1, x_2, x_3, x_4. If we substitute the values (230), we obtain a homogeneous quantity of the second degree in the increments dx'_a, thus

$$\sum\nolimits^{(ab)} g_{ab}dx_adx_b = \sum\nolimits^{(ab)} g'_{ab}dx'_adx'_b. \tag{244}$$

What is the relation between the two sets of quantities g_{ab} and g'_{ab}? When the substitution is made we have

$$\sum\nolimits^{(abce)} g_{ab}p_{ac}p_{be}dx'_cdx'_e.$$

Give definite values to c and e. The quantity $dx'_cdx'_e$ is then multiplied by

$$\sum\nolimits^{(ab)} p_{ac}p_{be}g_{ab},$$

and this must therefore be the value of g'_{ce}; that is,

$$g'_{ce} = \sum\nolimits^{(ab)} p_{ac}p_{be}g_{ab}\,; \tag{245}$$

g_{ab} is a tensor of the second rank, and (245) is its formula of transformation. There are sixteen equations of this form and sixteen terms on the right-hand side of each.

There is a difference between (245) and (242), for in (242) we have the coefficients π, whereas in (245) we have the

coefficients p, though in both cases the values in the new system are expressed in terms of the old. When we have the p's in the formulæ of transformation which express the new values in terms of the old ones, the tensor is said to be a *covariant*; when we have the π's, it is said to be a *contravariant*. It is to be remarked that when we invert the formulæ of transformation; that is, when we express the old values in terms of the new ones, we get p's where we had π's, and conversely.

Equation (235) shows that dx_a is a contravariant tensor.

It is often convenient to use upper suffixes in the components of a contravariant tensor and lower ones in the components of a covariant tensor, but we have written dx_a instead of dx^a for the reason already mentioned.

When $g_{ab} = g_{ba}$ we have also, by virtue of (245), $g'_{ce} = g'_{ec}$.

Proof. Interchange c and e in (245):

$$g'_{ec} = \sum^{(ab)} p_{ae} p_{bc} g_{ab} = \sum^{(ab)} p_{ae} p_{bc} g_{ba}.$$

But according to the transformation formula this is exactly the value of g'_{ce}, because interchanging a and b leaves the expression unaltered; thus

$$g'_{ec} = g'_{ce}. \tag{246}$$

When in a physical problem we have to deal with a quadratic homogeneous function of, say, three variables,

$$F = \tfrac{1}{2}(a_{11}x_1^2 + \cdots + 2\,a_{12}x_1x_2 + \cdots), \ (a_{21} = a_{12} \text{ etc.}) \tag{247}$$

the three partial derivatives ξ_1, ξ_2, ξ_3 are often interesting,

$$\xi_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3,$$
$$\xi_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3,$$
$$\xi_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3,$$

and we have the theorem

$$F = \tfrac{1}{2}(x_1\xi_1 + x_2\xi_2 + x_3\xi_3). \tag{248}$$

For instance, x_1, x_2, x_3 may be the component angular velocities of a body about the three axes, and ξ_1, ξ_2, ξ_3 the components of the moment of momentum. Or, again, x_1, x_2, x_3 may be the

components of the electric force in a crystal, and ξ_1, ξ_2, ξ_3 the components of the dielectric displacement. Having

$$a_{11}x_1^2 + \cdots + 2\,a_{12}x_1x_2 + \cdots,$$

we write it in the form

$$x_1\xi_1 + x_2\xi_2 + x_3\xi_3$$

by reckoning $2\,a_{12}x_1x_2$ as belonging half to the term with factor x_1 and half to that with factor x_2.

In the case now under consideration we can put

$$Q = \sum{}^{(ab)}\, g_{ab}dx_adx_b = \sum{}^{(a)}\, \xi_a dx_a, \tag{249}$$

where
$$\xi_a = \sum{}^{(b)}\, g_{ab}dx_b. \tag{250}$$

ξ_a is a covariant tensor.

Proof. We have

$$\xi'_a = \sum{}^{(b)}\, g_{ab}dx'_b = \sum{}^{(bcef)}\, p_{ca}p_{eb}\pi_{fb}g_{ce}dx_f.$$

Give to b the values 1, 2, 3, 4 with special values of c, e, f.

Then
$$\sum{}^{(b)}\, p_{eb}\pi_{fb}\begin{Bmatrix} =1 & \text{for} & f=e \\ =0 & \text{for} & f\neq e\end{Bmatrix},$$

and therefore
$$\xi'_a = \sum{}^{(ce)}\, p_{ca}g_{ce}dx_e = \sum{}^{(c)}\, p_{ca}\xi_c.$$

Inversely,
$$\xi_a = \sum{}^{(b)}\, \pi_{ab}\xi'_b. \tag{251}$$

From (250) we derive the values

$$dx_a = \sum{}^{(b)}\, g^{ab}\xi_b, \tag{252}$$

where the set of quantities g^{ab} is inverse to the set g_{ab}.

Now our sum Q can be written in the forms

$$\sum{}^{(a)}\, \xi_a dx_a \quad \text{and} \quad \sum{}^{(a)}\, \xi'_a dx'_a,$$

and therefore in the forms

$$\sum{}^{(ab)}\, g^{ab}\xi_a\xi_b \quad \text{and} \quad \sum{}^{(ab)}\, g'^{ab}\xi'_a\xi'_b,$$

so that these also must be equal. We conclude that

$$g'^{ce} = \sum{}^{(ab)}\, \pi_{ac}\pi_{be}g^{ab} \tag{253}$$

and hence that g^{ab} is a contravariant tensor of the second rank.

What we now have found about these tensors will suffice for our purpose. We shall give only one or two more theorems.

The definition can easily be extended to tensors of higher rank. Suppose that k_{abc} is a covariant tensor of the third rank, and l^{bc} a contravariant tensor of the second rank. The sum

$$\sum_{(bc)} k_{abc} l^{bc}$$

will be different according to the value of a. There are therefore four sums, m_1, m_2, m_3, m_4. There is now the theorem that these four quantities form a covariant tensor of the first rank, or a covariant vector. (The proper notation for this has already been chosen.)

Proof. $$m_a = \sum_{(bc)} k_{abc} l^{bc},$$

and similarly $$m'_a = \sum_{(bc)} k'_{abc} l'^{bc}.$$

Thus $$m'_a = \sum_{(bcefghi)} p_{ea} p_{fb} p_{gc} \pi_{hb} \pi_{ic} k_{efg} l^{hi}.$$

First take the sum

$$\sum_{(c)} p_{gc} \pi_{ic} \begin{cases} = 1 & \text{for} \quad i = g \\ = 0 & \text{for} \quad i \neq g \end{cases};$$

the expression for m'_a then reduces to

$$m'_a = \sum_{(befgh)} p_{ea} p_{fb} \pi_{hb} k_{efg} l^{hg}.$$

Next perform the summation

$$\sum_{(b)} p_{fb} \pi_{hb} \begin{cases} = 1 & \text{for} \quad h = f \\ = 0 & \text{for} \quad h \neq f \end{cases};$$

then $$m'_a = \sum_{(efg)} p_{ea} k_{efg} l^{fg} = \sum_{(e)} p_{ea} m_e.$$

The expressions $\sum_{(a)} k_a l^a$ and $\sum_{(bc)} k_{bc} l^{bc}$

are peculiarly simple; each of them is a single number that does not refer to one or more of the coördinates. This number is the same whether we use one set of coördinates or the other; it is an invariant or scalar quantity.

Einstein has succeeded (and this is a most remarkable achievement) in describing physical phenomena in a way that is the same in all systems of coördinates. This is due to the fact that in the system x_1, x_2, x_3, x_4 he makes everything depend on the quadratic form which we denoted provisionally by Q, with the relation $g_{ab} = g_{ba}$.

1. In the first place, the velocity with which light is propagated may be derived from the quadratic form Q. Let us follow a light-signal in its course. At each time x_4 it has reached a definite position x_1, x_2, x_3 and during the time dx_4 the three space coördinates will have changed by dx_1, dx_2, dx_3. Now according to Einstein's theory these values substituted in the quadratic expression

$$\sum_{(ab)} g_{ab} dx_a dx_b$$

will make it vanish. If we divide by dx_4, and if we put $\frac{dx_a}{dx_4} = v_a$, we get

$$\sum_{(ab)} g_{ab} v_a v_b = 0; \tag{254}$$

v_1, v_2, v_3 are the components (in the three-dimensional space R_3) of the velocity with which the light-signal is propagated, and $v_4 = 1$.

If the equation is written out, we have

$$g_{11} v_1^2 + \cdots + 2\, g_{12} v_1 v_2 + \cdots + 2\, g_{14} v_1 + \cdots + g_{44} = 0,$$

a formula that enables us to determine the velocity of light when the direction of the ray is given. The velocity in question is the ray velocity, and when its direction is known we know also the ratios between v_1, v_2, v_3 and v itself, so that (254) leads to an equation (of the second degree) for the determination of v.

2. We have good reasons to suppose that in parts of space far away from all bodies that produce a gravitational field we can introduce rectangular space coördinates x_1, x_2, x_3 and a time x_4 in such a manner that the velocity of light has the constant value c in all directions. Then equation (254) must reduce to

$$v^2 = c^2.$$

This means that in the present case we can ascribe to the g_{ab}'s the following values:

$$g_{11} = g_{22} = g_{33} = -1, \quad g_{44} = c^2, \quad g_{ab} = 0 \text{ for } a \neq b. \tag{255}$$

If the g's can be given these values, which we may call the "normal" ones, we shall say that the space is free from gravitation.

If, however, the coördinates cannot be so chosen that we have the values (255), we shall speak of a gravitational field, in

which the ten quantities g_{ab} are the "potentials." The reason why we call it a gravitational field will soon become clear.

3. Suppose that in the system x_1, x_2, x_3, x_4 we have the values (255); then by means of the transformation formula (245) we can determine the potentials g'_{ab} in the system x'_1, x'_2, x'_3, x'_4. In general they will be different from (255), and it looks much as if there were a gravitational field. It is, however, not a real but a spurious one. A real gravitational field is always connected with some attracting body; so that we cannot create it by a simple mathematical transformation. We may, however, if we like to do so, speak also of a gravitational field even when it is only apparent.

4. Consider the special relativity transformation

$$x'_1 = x_1, \quad x'_2 = x_2, \quad x'_3 = ax_3 - bcx_4, \quad x'_4 = ax_4 - \frac{b}{c}x_3 \quad (256)$$

with the relation $\qquad a^2 - b^2 = 1.$

These formulæ are easily inverted (as we know), so that the values of both the p_{ab}'s and π_{ab}'s are known. We can therefore, starting from (255), find g'_{ab} by (245). It is found that these still have the values (255), as we knew already.

5. In some cases the change in the velocity of light due to a transformation to new coördinates is a very simple matter. For instance, if we first have the values (255) and then put

$$x'_1 = kx_1, \quad x'_2 = kx_2, \quad x'_3 = kx_3, \quad x'_4 = lx_4,$$

k and l being constants, we find

$$g'_{11} = g'_{22} = g'_{33} = -\frac{1}{k^2}; \quad g'_{44} = \frac{c^2}{l^2}, \quad g'_{ab} = 0 \text{ for } a \neq b,$$

so that in the new coördinates (254) becomes

$$-\frac{v^2}{k^2} + \frac{c^2}{l^2} = 0,$$

$$v = \pm \frac{k}{l} c.$$

This is simply a change in the numerical value of the speed of light, which is due to changes in the units of length and time. The meaning of the double sign will be obvious.

6. Formula (254) leads to a quadratic equation for the velocity of light in a given direction; the roots of this equation are not equal and opposite when the coefficients g_{14}, g_{24}, g_{34} are different from zero. This can occur when, starting from a system in which values (255) hold, we make the substitution

$$x_1' = x_1 - wx_4, \quad x_2' = x_2, \quad x_3' = x_3, \quad x_4' = x_4,$$

where w is a constant. We then have

$$g_{11}' = g_{22}' = g_{33}' = -1, \quad g_{14}' = -w, \quad g_{44}' = c^2 - w^2,$$

so that if a ray has the direction of x_1, equation (254) becomes

$$-v_1^2 - 2\,wv_1 + c^2 - w^2 = 0,$$

giving
$$v_1 = \pm c - w,$$

a result that can easily be interpreted.

7. Whereas in this case nobody will even speak of a spurious gravitation field, (254) also applies in Einstein's theory to all real gravitational fields. The deflection of rays of light by the attraction of the sun can be deduced from it if we know the potentials.

We shall now consider the motion of a material particle. This also is determined by the quadratic form which we have already used. We shall henceforth put

$$ds^2 = \sum{}_{(ab)}\, g_{ab}dx_a dx_b \tag{257}$$

and shall call ds the line-element in the space R_4.

It may be recalled that there are many different geometries, —for instance that of the plane, that of the sphere, etc. We could also have a geometry of the surface of an ellipsoid. Now in all these geometries we can begin by choosing a proper number of coördinates and expressing the distance between two infinitely near points in terms of the differences of their coördinates. For example, with rectangular coördinates

$$ds^2 = dx_1^2 + dx_2^2 \text{ in a plane,}$$
$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 \text{ in space;}$$

with oblique coördinates the relations are more complicated, the product terms also appearing; with polar coördinates on a sphere of radius a (Fig. 44)

$$ds^2 = a^2 \sin^2 x_2 \cdot dx_1^2 + a^2 dx_2^2.$$

A similar expression can also be found in the case of an ellipsoid; ds^2 is always a homogeneous quadratic function of the differentials dx. If the formula for the line-element is known, the whole geometry can be based on it. We shall revert to this later; at present we shall merely observe that angles, change of direction, curvature of lines, magnitudes of areas, etc. can all be found.

So our formula for ds^2 is the starting-point for a geometry of the space R_4; and since this geometry depends on g_{ab}, it will be different in the case of a gravitational field than when there is none.

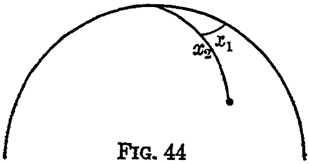

FIG. 44

The motion of a material particle is represented by its world-line, and therefore the law that regulates the motion must give us the course of the world-line. Einstein's law is as follows:

If P_1 and P_2 are two points in R_4, and L (going from P_1 to P_2) the world-line of a point, and moreover, if L' is a line differing infinitely little from L and between the same points, then the variation of the length when we pass from L to L' is zero; that is, it is a quantity of the second order. Thus

$$\delta \int ds = 0. \qquad (258)$$

The world-line is a *geodesic line* in R_4, but the length is not necessarily a minimum.

It is clear that we have now a way of describing motion that is the same in $x_1 \cdots x_4$ and $x_1' \cdots x_4'$. In the latter system

$$ds^2 = \sum_{(ab)} g_{ab}' dx_a' dx_b',$$

where g_{ab}' is related to g_{ab} in a way that we have already described. But in the new system, with the aid of the g_{ab}''s we

can describe phenomena just as in the old system we did with the g_{ab}'s. We thus have physical equations which are covariant.

The condition (258) may look somewhat strange at first, but in ordinary mechanics similar theorems occur. A point that is constrained to remain on a surface without friction will describe a geodesic line; and in all cases in which Hamilton's principle or the principle of least action applies, the real motion is distinguished from other conceivable motions by the condition that some definite integral remains stationary when the motion is varied to an infinitely small extent.

From the condition in Einstein's theory that the world-line of a particle is a geodesic in R_4 we can deduce the equations of motion in the form*

$$\frac{d}{dt}(-q_c) = \sum{}_{(ab)} \omega_{abc} \frac{\partial g_{ab}}{\partial x_c}. \tag{259}$$

($c = 1, 2, 3.$) In these equations $-q_1$, $-q_2$, $-q_3$ are the components of the momentum of the particle, and the expressions on the right-hand side, in which ω_{abc} are certain coefficients which we shall not further calculate, may be considered as the expressions for the component forces exerted on the particle by the gravitational field. We have three equations of this kind with $c = 1, 2, 3$.

The quantities q_a are defined by the formula

$$q_a = \frac{m}{ds} \sum{}_{(b)} g_{ab} dx_b. \tag{260}$$

It is remarkable that there is a fourth equation that is obtained by taking $c = 4$ in (259). This is the equation of energy, the energy of the particle being q_4, calculated according to (260). The second member of (259) taken for $c = 4$ relates to the amount of energy which the particle receives per unit of time in the gravitational field. The formulæ show that an influence of the field on the momentum and energy of the particle exists only when the potentials g_{ab} are not constants but are functions of the coördinates.*

* See Note 18, Appendix.

It is interesting to see what (260) becomes when the potentials have their normal values (255). Then we have

$$\left.\begin{aligned} q_a = mg_{aa}\frac{dx_a}{ds} \\ ds^2 = (c^2 - v^2)dx_4^2, \quad ds = \sqrt{c^2 - v^2}\,dx_4 \end{aligned}\right\}, \tag{261}$$

so that the momenta become

$$-q_1 = \frac{mv_1}{\sqrt{c^2 - v^2}}, \quad -q_2 = \frac{mv_2}{\sqrt{c^2 - v^2}}, \quad -q_3 = \frac{mv_3}{\sqrt{c^2 - v^2}}, \tag{262}$$

and the energy becomes $\dfrac{mc^2}{\sqrt{c^2 - v^2}}$. (263)

These are the values which we found in special relativity, only divided by a factor c. We can, however, always replace m by cm, and then we have exactly the old values.

As (260) shows, q_a is a covariant vector.

We can now, just as we did in special relativity, consider a system composed of a great number of points all (at least all that are in an element of volume) moving with the same velocity with components (v_1, v_2, v_3). At a definite instant x_4 let $Ndx_1dx_2dx_3$ be the number of particles in the element dx_1, dx_2, dx_3, so that N may be called the number per unit of volume. Then, for instance, the energy per unit of volume is

$$Nq_4, \tag{264}$$

and the first component of momentum per unit of volume is

$$-Nq_1. \tag{265}$$

The first component of the flow of energy is

$$Nq_4v_1. \tag{266}$$

We can further consider an element dx_2dx_3 and fix our attention on the momentum $-q_1$ that passes through it per unit of time. So we are led to the first of the stress components, for which we may still use the ordinary expressions X_x etc. if we write x, y, z instead of x_1, x_2, x_3. The result is

$$X_x = Nq_1v_1. \tag{267}$$

Indeed, $X_xdy\,dz$ means the change per unit of time in the momentum in the direction of x, existing on the negative side

of the element $dy\,dz$ in so far as that change is due to what happens at the element. In our case the only cause of the change is the transfer of the momentum by particles going through the element $dy\,dz$.

The above expressions can be simplified if, in addition to v_1, v_2, v_3, we introduce

$$v_4 = \frac{dx_4}{dt} = 1, \tag{268}$$

multiplying by this the expressions in which there is as yet no factor representing a velocity. We can then say that if $T_1^1, T_1^2, \cdots$ are the stresses $X_x, X_y, \cdots$, and $-T_1^4 \cdots$ the momenta per unit of volume, $T_4^1 \cdots$ the components of the flow of energy, and T_4^4 the energy itself per unit of volume, we have

$$T_a^b = N q_a v_b. \tag{269}$$

From this we can deduce the transformation formulæ, not forgetting that when we pass to $x_1' \cdots x_4'$ we shall have not only another q_a and another v_b (namely, q_a' and v_b') but also another N; namely, N'.

In examining the relation between N and N' we are led to introduce the determinant of the coefficients p_{ab}, which is equal to

$$\frac{\sqrt{-g'}}{\sqrt{-g}}.$$

Here g is the determinant of the potentials g_{ab}, and, similarly g' is the determinant of the potentials g'_{ab}. Instead of these quantities we have introduced $-g$ and $-g'$, because g and g' are always negative. This is seen immediately when the g_{ab}'s have their normal values, for then $g = -c^2$.

It is easily found that

$$\frac{1}{\sqrt{-g}} T_a^b$$

is a tensor, covariant with respect to the suffix a and contravariant with respect to b ("mixed" tensor). Thus*

$$\frac{1}{\sqrt{-g'}} T_a'^b = \frac{1}{\sqrt{-g}} \sum{}^{(ce)} p_{ca} \pi_{eb} T_c^e. \tag{270}$$

* See Note 19, Appendix.

From the mixed tensor $\frac{1}{\sqrt{-g}} T_a^b$ we can deduce a covariant tensor of the second rank T_{ab} by means of the formula

$$T_{ab} = \sum^{(c)} \frac{1}{\sqrt{-g}} g_{cb} T_a^c. \tag{271}$$

This tensor is symmetrical;* that is, $T_{ba} = T_{ab}$, as can be deduced from (269) combined with the value of q_a.

In all other cases also the components of momentum etc. are considered to be such that

$$\frac{1}{\sqrt{-g}} T_a^b$$

is a mixed tensor and that T_{ab} is symmetrical. This can be verified for the case of the electromagnetic field; the equations for this field must, however, be modified in a way that Einstein has indicated. In connection with these equations the formulæ for the momenta etc. can be found.

It may be remarked that Maxwell's equations in their generalized form enable us to answer all questions concerning the propagation of light in a gravitational field. In the first place, we can deduce from them the law of propagation of waves; then, by Huygens's construction, the same for rays. It is found that the velocity of the rays is determined by $ds^2 = 0$.

59. Einstein's Field Equations. In the old theory of gravitation there is only one potential, and the gravitation is simply due to the mass of the attracting body. The distribution being determined by the density ρ, we have for the potential Poisson's equation

$$\Delta\phi = 4\,\pi\kappa\rho,$$

where κ is the constant of gravitation and

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

This is a partial differential equation of the second order for the determination of ϕ when ρ is given.

* See Note 20, Appendix.

Now in Einstein's theory the ten potentials g_{ab} have to be determined by partial differential equations as functions of the coördinates, and this has to be done by means of formulæ that do not change form when new coördinates are chosen. It was certainly a most remarkable achievement to obtain equations of this kind.

We have already seen that our argument that

$$ds^2 = \sum_{(ab)} g_{ab}dx_a dx_b$$

gives us the length of the line-element in the space R_4 is the starting-point of what we may call the geometry of R_4. The geometry can be fully developed when we know the g_{ab}'s as functions of the coördinates. We can then deduce differential equations for the geodesic lines, and can also determine angles.

If we wish to determine the angle between the lines AB and AC, we take along each an indefinitely short step and join the ends B, C. We then calculate by our fundamental formula the lengths of AB, AC, BC and deduce from them the cosine of the angle A just as if the triangle were an ordinary plane triangle.

We can also define what is meant by an infinitely small plane passing through AB and AC. In much of what has been said thus far it has been tacitly assumed that in our four-dimensional space two points that are infinitely near each other can be joined by what we may call a straight line-element, the simultaneous changes of the four coördinates being in constant ratios to each other when we pass along the line. The plane in question is determined by all the line-elements joining A and points of BC, all these lines lying in the plane.

At a definite point of the space R_4 we have many different planes; for instance, 1, 2; 1, 3; 1, 4; etc., the meaning of these notations being as follows: A line-element along which, of the four coördinates, x_1 only changes may be called 1, and line-elements 2, 3, 4 may be defined in a similar way. The symbol 1, 2 represents the plane passing through 1 and 2.

Let AB be an infinitely small displacement or vector of length ds and components dx_a. We shall call the quantities

$$\eta^a = \frac{dx_a}{ds} \tag{272}$$

its direction constants. It is clear that they satisfy the condition

$$\sum\nolimits^{(ab)} g_{ab}\eta^a\eta^b = 1. \tag{273}$$

We must now consider the following important problem: Suppose that at the point P (Fig. 45) we have a vector PA with the direction constants η^a. We displace P along a geodesic line PL. We can imagine that in doing so we cause the vector PA to move along with the point P, its direction constants being gradually changed according to some rule agreed upon.

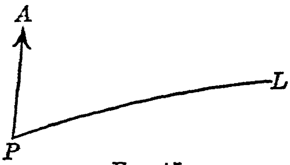

FIG. 45

Equation (273) shows that they cannot in general be kept constant. Now this rule can be chosen in such a way that we get the nearest approach to what we may call a displacement of PA parallel to itself. For this purpose we may proceed in the following way:*

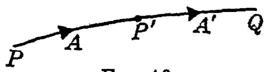

FIG. 46

1. Suppose (Fig. 46) the vector PA to have the direction of the geodesic line $PP'Q$ along which it is to be displaced. Then it must constantly fall along that line. This amounts to saying that a second element of a geodesic line is parallel to the first.

2. Let PA be at right angles to the geodesic line PL (Fig. 47), and let P' be infinitely close to P. Then it is natural to assume that if PA remains parallel to itself, it must be also perpendicular to PL at the point P'. But its direction is not determined by this condition, because at P' the number

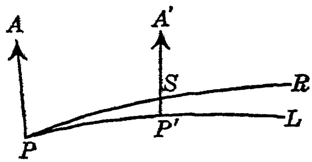

FIG. 47

* For the theory of parallelism reference may be made to T. Levi Civita, *Rend. Palermo*, Vol. 42 (1917), p. 1.

of directions perpendicular to a given line is infinite. The direction $P'A'$ can, however, be completely determined as follows:

Let PR be a second geodesic line differing infinitely little from PL and having at P an initial direction which lies in a plane containing the initial directions of PL and PA. Take on PR a length PS equal to PP'; then the line $P'S$ will give the required direction.

3. Let PA (Fig. 48) have any direction. Choose PB so that it is perpendicular to PL and so that at P the three directions are in one plane. Let PB be displaced along PL in the way just explained, and let $P'B'$ be a new position. Then the direction $P'A'$ which we want to determine is at P' in the same plane with PL and $P'B'$, and it is further determined by the condition that it must make the same angle with the geodesic line $P'L$ as PA makes with PL.

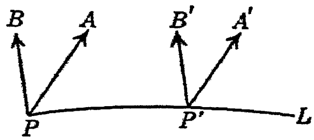

FIG. 48

This definition of a vector's displacement parallel to itself can be applied to ordinary geometry and also to the geometry of the sphere and ellipsoid. We are then led to remarkable consequences. Suppose the point P is made to move along some closed curve or polygon, so that we come back to the starting-point P_0. Then if a vector accompanies the point in this motion, so that each successive position of it, in the above sense of the word, is parallel to the preceding one, the vector will just come back to its initial direction if we have to do with a plane or an ordinary space of three dimensions, the Euclidean space. But things are different in the geometry of the sphere. According to our argument a vector on a sphere, when displaced along a great circle parallel to itself, will simply make the same angle with the circle all the time. Now let us apply this to a displacement of the vector 0 (Fig. 49) along the sides of the spherical triangle ABC, so that it will take successively the positions 1, 2, 3. The last of these will not coincide with 0 (Fig. 49).

Indeed at B we have the angle θ between 1 and AB produced. Thus the angle between 1 and BC is $180° - B - \theta$; the angle

between 2 and CB is $B+\theta$, and the angle between 2 and AC prolonged is $B+\theta-(180°-C)=B+C+\theta-180°$. This is also the angle between 3 and AC, so that the angle between 3 and AB is $A+B+C+\theta-180°$. Therefore 3 makes with 0 an angle $(A+B+C)-180°$, which, as you know, is different from zero. The difference is proportional to the area of the spherical triangle. The same result in a general form holds also in our R_4.

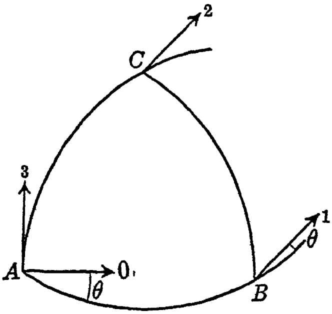

FIG. 49

The above property of the sphere arises, as we say, from the curvature of the surface. Similarly, we may speak of the curvature of the space R_4. It possesses even curvatures in different directions, for at any point we can distinguish between the planes 1, 2; 1, 3; etc., and we can consider closed lines in each of them.

Let us take two infinitely short lines PQ and PR (Fig. 50) having the directions of x_k and x_l, and let dx_k and dx_l be the changes of x_k and x_l when we go from P to Q or from P to R.

After completing the parallelogram $PQSR$ (this means that for each of the coördinates the change when we go from Q to S is equal to the change belonging to the displacement PR) let us move along this circuit in the direction $PQSRP$ (dx_k first). Then if η^1, η^2, η^3, η^4 are the direction constants of some vector at P at the start, and if this vector accompanies the motion along the circuit, remaining parallel to itself, the direction constants η^a will be changed when we come back to the starting-point P by

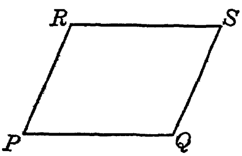

FIG. 50

$$\delta\eta^a=\sum_{(c)} H^{\;a}_{ckl}\,\eta^c dx_k dx_l, \tag{274}$$

where $H^{\;a}_{ckl}$ is an expression containing the first and the second derivatives of g_{ab} with respect to the coördinates. It can be wholly developed, but we shall omit the formula. For our

purpose it is sufficient to remark that H^a_{ckl} is a mixed tensor of the fourth rank, partly covariant and partly contravariant, as has been indicated by the position of the suffixes.*

From H^a_{ckl} we can deduce a covariant tensor of the second rank; namely,

$$G_{ck} = \sum\nolimits_{(a)} H^a_{cka}. \tag{275}$$

H^a_{cka} is what H^a_{ckl} becomes when $l = a$; that is, when we want to know the change of the direction constant with respect to a coördinate whose increment is represented by one of the sides of the parallelogram.

That G_{ck} really is a covariant tensor can be easily proved from (275). Moreover, when the value of H^a_{cka} is substituted, it is found that G_{ck} is symmetrical;† that is, $G_{ck} = G_{kc}$.

Now Einstein's field equations have the form

$$G_{ab} = -\kappa\,(T_{ab} - \tfrac{1}{2}\,g_{ab}T). \tag{276}$$

Here T_{ab} is the covariant tensor which we found when speaking of the momenta, stresses, etc. T is the scalar defined by

$$T = \sum\nolimits_{(cd)} g^{cd}T_{cd}. \tag{277}$$

κ is the constant of gravitation; such a constant, of course, has to be introduced.

There are ten equations of this form which partially determine the ten potentials g_{ab}.

Later on, Einstein changed his equations, giving them the form

$$G_{ab} - \lambda g_{ab} = -\kappa(T_{ab} - \tfrac{1}{2}\,g_{ab}T), \tag{278}$$

where λ is a new constant.

Both the equations (276) and (278) have the same form in all systems of coördinates. This means that if in the system x_1, x_2, x_3, x_4 we have equations (276), we shall have in x'_1, x'_2, x'_3, x'_4

$$G'_{ab} = -\kappa(T'_{ab} - \tfrac{1}{2}\,g'_{ab}T'),$$

where T'_{ab} is a tensor connected with T_{ab} by the transformation formulæ which we have already found, while T' is defined by a formula similar to (277) and G'_{ab} is made up of differential coefficients of g'_{ab} with respect to $x'_1 \cdots x'_4$ in exactly the same

* See Note 21, Appendix. † See Note 22, Appendix.

way as G_{ab} is made up of the derivatives of g_{ab} with respect to $x_1 \cdots x_4$. That the equations are true for any choice of coördinates is due to the fact that all terms are tensors of the same kind.

The gravitational field depends on the energy, the stresses, the momenta, and the flow of energy. They all tend to produce it, but the most important of them all is the energy. It replaces the mass in the old theory. You will remember that already in special relativity, when we have a mass m that is at rest, we have an energy mc^2.

Just as the energy T_4^4 is the most important of all the components of the tensor T_a^b, so g_{44} is the most important of the gravitational potentials, and by this it is brought about that, after all, we do not get far away from Newton's theory. But all this is due to the fact that velocities of bodies are always very small in comparison with that of light.

The influence of the other components of T_a^b on the gravitational field, though too small to be observed, is most interesting from a theoretical point of view. For instance, we can consider the sun in a first approximation as a globe of energy, the value of T_4^4 having in its interior a certain constant value; but if we wanted to go farther, we should have to take into account the distribution of the energy, which will not be exactly uniform, and also the internal stresses and pressures. If the sun is rotating, we shall have a momentum and a flow of energy; these also will contribute to the gravitational field. Even a beam of light will produce a field of that kind, so that if we have two beams parallel to each other and at a certain distance, the one may be deflected by the gravitational field that is produced by the other.

A cavity filled with black radiation will produce a gravitational field of its own; that is, it will attract bodies, say, more than when the radiation is not there. And here I may remark that it will also be attracted in a degree that is proportional to its attractive power and so to what we have called its mass.

Thus one of the most important features of the new theory is that gravitation is no longer an isolated phenomenon but is

closely connected with all other physical phenomena, — not, however, on an equal footing with them. It is a phenomenon of a more general kind and in a sense embraces them all.

If the theory can be maintained, this will certainly be one of its greatest beauties, and it is surely one of the greatest achievements in physical science to have established a doctrine so far-reaching and all-embracing.

I may add that the new theory throws light on things with which we have long been familiar. One of these is that with the same initial conditions all bodies fall so that at any instant they have the same velocity. Of course this has always attracted the attention of philosophers; it seemed to indicate that all bodies have something in common in their constitution. In Einstein's theory it is simply a consequence of the rule that the world-lines of all material particles or bodies (in so far as these can be considered as particles, not too large) are geodesic lines.

It may here be recalled that the expressions which we found in special relativity for the momenta and energy of a particle can, as we showed, also be applied to any system in a stationary state. The same is true in the gravitational field, provided the dimensions are small compared with those of the field. Also, the changes in the momenta that are due to the gradients of the potentials are determined in the same way as in the case of a material particle, so that the equations of motion remain the same. If we had black radiation inclosed in an envelope, — say in an envelope without mass or weight, — that black radiation would fall with the same velocity as a stone; its world-line would again be a geodesic line.

Thus black radiation, and so also a beam of light, has really weight. This weight can be transmitted to the cavity in the same way as the weight of a mass of air, the mechanism of the transmission being that the rays of light (curved by the gravitational field) exert pressures on the walls that have a definite resultant.

Let us revert from these imaginary cases to phenomena which we observe daily. In the ordinary gravitational field of the earth, which is uniform over not too great an extent, the

potential g_{44} is a linear function of the height. For bodies moving with small velocities the momenta are mv_1, mv_2, mv_3. The rate of change of the first of these, if x_1 is vertical, is connected with the gradient $\frac{\partial g_{44}}{\partial x_1}$ by our first equation of motion, all as in the old theory.

I may here remark that, just as in this case, when we want to solve a definite problem we do not use strange or quixotic coördinates but coördinates by which everything may become as simple as possible.

60. Some Further Observations on Einstein's Theory. 1. Einstein's field equations (276) have been solved for several special cases, so that we are able to give the values of the potentials; that is, of the coefficients in the formula for ds^2. The simplest case is that of the field around a body with spherical symmetry (or around any body if we remain far enough away from it). Schwarzschild has given an expression for ds^2 which takes the following form when we confine ourselves to a plane passing through the center; r and ϕ are ordinary polar coördinates in this plane.

$$ds^2 = -\frac{dr^2}{1-\frac{\alpha}{r}} - r^2 d\phi^2 + c^2\left(1-\frac{\alpha}{r}\right)dt^2. \tag{279}$$

As an example of the applications of this formula we shall deduce from it the conditions under which a circular orbit can be described with constant velocity. In this case $dr = 0$ and $d\phi = \omega\, dt$, if ω is the angular velocity. Further, in applying the fundamental formula (258), we may suppose only r to be varied, ϕ and t remaining unaltered. As the variation of dr^2, that is, $2\,dr\,\partial dr$, vanishes on account of the first factor, we find from (279)

$$2\,ds\,\delta ds = -2\,r\,\delta r\,d\phi^2 + \frac{\alpha c^2}{r^2}\,\delta r\,dt^2 = \left(-2\,r\omega^2 + \frac{\alpha c^2}{r^2}\right)\delta r\,dt^2.$$

This leads to the condition

$$-2\,r\omega^2 + \frac{\alpha c^2}{r^2} = 0, \quad \omega^2 = \frac{\alpha c^2}{2\,r^3},$$

agreeing with Kepler's third law.

Let us now introduce in (279) a new variable r', where $r = f(r')$; then

$$ds^2 = -\frac{[f'(r')]^2 dr'^2}{1 - \dfrac{\alpha}{f(r')}} - [f(r')]^2 d\phi^2 + c^2\left[1 - \frac{\alpha}{f(r')}\right] dt^2. \quad (280)$$

In particular, if we put $r = \dfrac{1}{r'}\left(r' + \dfrac{\alpha}{4}\right)^2$, (281)

we find that

$$ds^2 = -\left(1 + \frac{\alpha}{4\,r'}\right)^4 (dr'^2 + r'^2 d\phi^2) + c^2 \left(\frac{1 - \dfrac{\alpha}{4\,r'}}{1 + \dfrac{\alpha}{4\,r'}}\right)^2 dt^2. \quad (282)$$

2. The field equations have the same form whatever be our choice of coördinates; it does not matter whether we take as the first polar coördinate r or r'. Therefore the coefficients in (282), which satisfy the equations with r' as independent variable, will also, if we simply omit the primes, satisfy the original equations of which the coefficients of (279) formed a solution. So it is found that the field equations admit of different solutions. Why does this make no difference?

If we choose first one system of g_{ab}'s satisfying the field equations, and then another, we shall get different orbits and different courses for light-signals, but observable phenomena remain the same. All observable phenomena are in fact what we may call coincidences. A vibration of light, for instance, that has been emitted by a star coincides at a certain moment with a point of the moon's border, and at a later instant with the point of intersection of the cross wires in a telescope. It is clear that when there is a coincidence of this kind, two world-lines will have a point in common, which in general they have not.

Now suppose that if we take r, ϕ, and t as the independent variables, determining the propagation of light and the motion of bodies on the basis of the expression (279), we find two world-lines L_1 and L_2 whose equations are

$$\psi_1(r, \phi, t) = 0, \quad \chi_1(r, \phi, t) = 0, \quad (283)$$

and

$$\psi_2(r, \phi, t) = 0, \quad \chi_2(r, \phi, t) = 0, \quad (284)$$

and that these intersect at the point

$$r = \lambda, \quad \phi = \mu, \quad t = \nu, \quad (285)$$

these values satisfying the four equations.

If in (283) and (284) we substitute for r the expression (281), we find equations in r', ϕ, t of a somewhat different form, say

$$\bar{\psi}_1(r', \phi, t) = 0, \quad \bar{\chi}_1(r', \phi, t) = 0, \tag{286}$$

and

$$\bar{\psi}_2(r', \phi, t) = 0, \quad \bar{\chi}_2(r', \phi, t) = 0; \tag{287}$$

and these will have the common solution

$$r' = \bar{\lambda}, \quad \phi = \mu, \quad t = \nu, \tag{288}$$

$\bar{\lambda}$ being the value of r' that corresponds to $r = \lambda$.

It will be clear that if we keep to r as independent variable, but use the second solution of the field equations of which we have spoken, namely, the solution given by (282) with the dashes omitted, there will be world-lines whose equations are

$$\bar{\psi}_1(r, \phi, t) = 0, \quad \bar{\chi}_1(r, \phi, t) = 0,$$

and

$$\bar{\psi}_2(r, \phi, t) = 0, \quad \bar{\chi}_2(r, \phi, t) = 0.$$

These will be different from the lines (283) and (284), but like these they will intersect each other, the coördinates of their common point being

$$r = \bar{\lambda}, \quad \phi = \mu, \quad t = \nu.$$

3. Though the principle of relativity leaves us entirely free in the choice of coördinates, it will be advisable, in any particular case, to choose them so that the mathematical calculations and results take a form as simple as may be. It is true that according to Einstein's views we cannot think of three rectangular axes of coördinates fixed in the ether, nor of the time as measured by a particular system of clocks that occupy stationary positions with respect to this medium; but we can always try, by a proper choice of coördinates, to make the values of the potentials g_{ab} as simple as possible; we can often do so by using rectangular space coördinates x, y, z. If in this case we define our line-element by

$$d\sigma^2 = dx^2 + dy^2 + dz^2, \tag{289}$$

we have ordinary geometry and no curvature of space; but if, in the case of a gravitational field, confining ourselves to $x_4 = \text{const.}$, $dx_4 = 0$, we define the line-element by

$$ds^2 = \sum_{(a,\, b = 1,\, 2,\, 3)} g_{ab} dx_a dx_b,$$

space will be curved. This curvature can be detected by means of measuring rods, just as we can detect that of a spherical surface, for instance.

Suppose we have a very great number of equal rods, so short that they may be said to fit to the surface. On a plane we can form with these rods a regular hexagon, each side consisting of n rods laid end to end, and the six radii of which are made up of n rods each. But if n is large enough, this will be found impossible in the case of the sphere. If we were endowed with unlimited means of observation, we should see that a similar experiment performed in a gravitational field fails likewise, and this would show us what we have called the curvature of the field. Nevertheless we can also speak, if we like, in terms of the geometry characterized by (289). If we prefer this, our space is not curved, and the failure of the hexagon experiment is to be ascribed to small changes in the lengths of the rods caused by the gravitational field.*

61. Displacement of the Lines in the Solar Spectrum toward the Red. We can imagine an instrument of very small size in which a periodic motion is continually going on, so that the "clock," as we may call it, gives a regular succession of ticks. According to the theory of relativity it must be possible to describe the fact of this regularity in terms that are independent of the choice of coördinates. This would not be the case if it consisted in an equality of time intervals, for when the coördinate x_4 changes by equal steps, the coördinate x'_4 will generally not do so. It is always possible, however, to think of the world-line L of the clock and of the points P, P', P'', $\cdots$ of that line, corresponding to the ticks. Supposing these points to be infinitely near each other, we shall be true to the principle of relativity if we assume that in the case of a "perfect" clock the elements PP', $P'P''$, etc. of the world-line have always one and the same length when we calculate it by means of our formula for ds^2. Thus

$$\sum_{(ab)} g_{ab} dx_a dx_b = l^2,$$

* See Note 23, Appendix.

where l has a value characteristic for the clock, the same, of course, for clocks of equal construction.

1. If the potentials g_{ab} have their normal values (255) the formula becomes (if v is the velocity with which the clock moves)

$$(c^2 - v^2)dx_4^2 = l^2,$$

$$dx_4 = \frac{l}{\sqrt{c^2 - v^2}},$$

showing, because x_4 is the "time," that the translation of the clock with a velocity v makes it move slower.

2. Let us next suppose that the clock is at rest in a gravitational field. Then

$$g_{44}dx_4^2 = l^2, \quad dx_4 = \frac{l}{\sqrt{g_{44}}}. \tag{290}$$

If, for instance, g_{44} has the value occurring in (279), the time interval between two ticks will be inversely proportional to $\sqrt{1 - \frac{\alpha}{r}}$. On account of the gravitation exerted by the sun, dx_4 will be greater for a clock that is placed at the sun's surface than for one that is placed on the earth. This is the reasoning that has led Einstein to his celebrated prediction that the lines in the solar spectrum, when compared with the corresponding lines of a terrestrial source of light, will be found to be displaced a little toward the red. The shift would be very slight, and, as you know, the reality of the phenomenon is still somewhat doubtful.*

REMARK 1. We can imagine a clock first to have a stationary position at the surface of the earth, and then to be transferred to the sun where it again takes a stationary position. The world-line L, which has a continuous course, will be in the direction x_4 both before and after the displacement, and we can apply (290) to two intervals lying in these extreme parts of the line.

Similarly, we can say that the period of its vibrations would be changed if an atom of sodium were transferred from the earth to the sun, and the conclusion may be applied to two atoms A and B, one on the earth and one on the sun, if we suppose that A, when transferred to the sun, would find its exact counterpart in B.

* Since this was said, the shift has been observed in the case of a star for which it is greater than for the sun.

REMARK 2. In what precedes we spoke of the lengths of the intervals dx_4 or dt corresponding to the succeeding vibrations. But how about the *observed* frequency?

The gravitational field is stationary, and in this case the electromagnetic equations admit solutions in which all expressions contain one and the same factor $e^{\iota nt}$, so that the frequency is everywhere equal to that of the source. We can see this without knowing the equations, because the conditions for propagation are the same for all succeeding vibrations.*

62. A Problem relating to the Effect of Centrifugal Force. It has been shown by the last eclipse expedition that light is deflected by gravitation. The question once arose as to whether rays of light can be similarly deflected by centrifugal force. I was asked, What will occur when, for instance, a glass cylinder is rotated about its axis and a ray of light is passed through it in the direction of the axis? Will it be deviated from the straight line? The answer is as follows:

First choose axes of coördinates that are fixed in the room, not in the rotating cylinder. What are the values of the potentials g_{ab}? If there were no gravitational fields, we should have the normal values of the potentials. Now there is the field of the earth, but we can neglect this because we want to know the influence of rotation only. Still more can we neglect the field produced by the glass cylinder and the change in it due to the rotation. Accordingly we have the normal values of the g_{ab}'s. This means that the phenomenon can be treated by means of the old theory and that it has nothing to do with general relativity.

We have only to take into account the fact that the glass, when rotating, drags along with it light-waves to an amount determined by Fresnel's dragging coefficient. It is found that there will really be some deflection of a ray of light in a plane passing through the axis. Needless to say, this deflection is much too small to be observed. The influence of the strains produced by centrifugal force and of the ensuing double refraction would be much greater.

* See Note 24, Appendix.

63. The Rotation of the Earth. This rotation, which has been established by Foucault's pendulum experiment and by the phenomena of falling bodies, can also conceivably be proved by electromagnetic or optical phenomena. Diurnal aberration has not been observed with certainty, but an interference experiment has been planned by Professor Michelson in which light is to be sent by means of reflecting mirrors round a triangle in two opposite directions, in order to see whether any difference in the time of propagation is produced by the earth's rotation.*

Consider an ideal experiment in which electric waves travel around two parallel wires, as in Lecher's system, except that the wires are supposed to be placed round the earth's equator, each forming a closed circle. The theory of relativity, as well as the old theory of a stationary ether, would require that the velocities of propagation in the two directions relative to the earth be unequal. There will then be standing waves whose nodes and loops move round the earth, making a complete circuit in a sidereal day. The earth may be said to rotate relative to these nodes and loops. (In this discussion we need not speak of the gravitational field of the earth; the phenomena would be the same if the earth exerted no gravitation at all.)

Now choose a system of coördinates. We can do this in two ways:

1. So that the loops and nodes are stationary; that is, so that the velocity of electric waves is the same in both directions. In this system we have the normal values of the potentials g_{ab}, and the earth appears to rotate.

2. So that the axes are fixed to the earth. We now have different values of the g_{ab}'s that can be found by means of the transformation formulæ. The new g_{ab}'s also occur in the equations of electromagnetism, and these show that the nodes and loops are rotating and that the velocities in the two directions are unequal.

Thus when we are asked *with respect to what* the earth rotates, we may say that it is rotating in a system of coördinates in

* Since this was written, the experiment has been performed, with the expected result.

which the g_{ab}'s have their normal values. Foucault's experiment leads to the same conclusion.

But we may wish to think of the earth's motion as motion relative to something that is more substantial than a mere system of axes of coördinates. Here there are two alternative views.

Physicists of former days would say that the nodes and loops have their seat in the ether, — that the earth rotates relative to this medium and that the g_{ab}'s have their normal values in a system of axes that is at rest in the ether. We have no longer to account for these normal values, the properties (if we may call them so) of the ether being responsible for them.

On the other hand, according to Einstein there is no such thing as the ether of which we have just spoken. If there is some medium, it has not sufficient substantiality to enable us to use it as a framework of reference with respect to which the position of bodies can be determined. Einstein thinks that in our experiment the nodes and loops are fixed in some way or other relative to the fixed stars, seen or perhaps unseen. This conception can be admitted because the stars are really seen rotating around the earth. Therefore we can admit that the nodes and loops are kept in their places by them, so that it is relative to them that the earth rotates.

Of course this implies that there is some kind of connection, or link, between the stars and the earth. Indeed, if there is nothing of the kind between them, the conception of relative motion is not wholly clear. We cannot, without further explanation, even draw a straight line toward some star. Now, according to Einstein, the system of coördinates in which the g_{ab}'s have their normal values is not determined by the fixed stars by some mysterious influence but by the gravitational field which they produce. In his opinion, even the normal values of the g_{ab}'s constitute a gravitational field due to the stars; and if the stars were not there, we should not have these normal values. This is connected with Einstein's idea that the inertia of a body must be considered as relative to something else; that is, to some other body. As a measure of inertia we can take momentum. The first component of the momentum is

given approximately by $-mg_{11}\dot{x}_1$; and when the g_{ab}'s have their normal values, this becomes $m\dot{x}_1$. Einstein thinks that if the stars were not there, we ought to have $g_{11}=0$, so that, even if the body were moving with respect to certain axes of coördinates, it would have no momentum.

I have now partially, though rather imperfectly, indicated the line of thought which led Einstein to his field equations in their later form (cf. equation (278)). I cannot enter into the solutions which they admit; it will be sufficient to say that he considers three-dimensional space as finite. The time coördinate x_4 can have all values ranging from $-\infty$ to $+\infty$, but the space coördinates are limited, so that the extension that corresponds to them is in three dimensions what the circumference of a circle is in one and the surface of a sphere in two. Like these latter extensions it has a definite radius R (immensely greater than any distance with which we are familiar), and the constant λ is connected with this radius.* Now, on account of the term with λ in the equation,

$$G_{ab}-\lambda g_{ab}=-\kappa(T_{ab}-\tfrac{1}{2}\,g_{ab}T),$$

the stars really determine the g_{ab}'s.

If we pass from the system of coördinates in which we have the normal g_{ab}'s to another fixed to the earth, we have different g_{ab}'s; these new values are again due to the stars, which are rotating in the new scheme and therefore produce a different gravitational field.

In performing the necessary calculation Einstein supposes the mass of the stars to be uniformly distributed, but we shall not speak of this.

64. The Time Variable in the Theory of Relativity. Let us finally revert for a few minutes to the special theory of relativity and to the transformation used in it, in which time also is involved,

$$x'=x,\quad y'=y,\quad z'=az-bct,\quad t'=at-\frac{b}{c}z.$$

You will remember our two observers A and B, using the different times t and t', and each able to describe physical

* See Note 25, Appendix.

phenomena in exactly the same way, though what is simultaneous for one is not simultaneous for the other. The theory of relativity emphasizes the fact that one of these is exactly as good as the other. A physicist of the old school says, "I prefer the time that is measured by a clock that is stationary in the ether, and I consider this as the true time, though I admit that I cannot make out which of the two times is the right one, that of A or that of B." The relativist, however, maintains that there cannot be the least question of one time being better than the other.

Of course this is a subject that we might discuss for a long time. Let me say only this: All our theories help us to form pictures, or images, of the world around us, and we try to do this in such a way that the phenomena may be coördinated as well as possible, and that we may see clearly the way in which they are connected. Now in forming these images we can use the notions of space and time that have always been familiar to us, and which I, for my part, consider as perfectly clear and, moreover, as distinct from one another. My notion of time is so definite that I clearly distinguish in my picture what is simultaneous and what is not.

The fact that physical phenomena can just as well be described in terms of z and t as in terms of z' and t' simply means that I can form my picture in two different ways; namely, by taking t or t' for my time. The principle of relativity teaches us that one of the two modes of description is just as good as the other. There is nothing very strange or inconceivable in this.

As to the ether (to return to it once more), though the conception of it has certain advantages, it must be admitted that if Einstein had maintained it he certainly would not have given us his theory, and so we are very grateful to him for not having gone along the old-fashioned roads.

APPENDIX

Note 1, § 4

Assuming

$$(H_x, H_y, H_z) = (p, q, r)\, e^{in\left(t + k - \frac{x\cos\alpha + y\cos\beta + z\cos\gamma}{v}\right)}$$

and making use of the relations $\mathbf{B} = \mathbf{H}$, $D_x = \epsilon_1 E_x$, $D_y = \epsilon_2 E_y$, $D_z = \epsilon_3 E_z$, we find on substituting in the field equations that

$$r \cos\beta - q \cos\gamma + \frac{v}{c}\epsilon_1 f = 0, \text{ etc.} \quad h \cos\beta - g \cos\gamma - \frac{v}{c} p = 0, \text{ etc.}$$

The elimination of p, q, r leaves us with the three equations

$$\left(\frac{v^2}{c^2}\epsilon_1 - 1\right) f = -\cos\alpha\,(f\cos\alpha + g\cos\beta + h\cos\gamma), \text{ etc.}$$

use having been made of the relation $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$.

Eliminating f, g, h, we find that

$$\frac{\cos^2\alpha}{\frac{v^2}{c^2}\epsilon_1 - 1} + \frac{\cos^2\beta}{\frac{v^2}{c^2}\epsilon_2 - 1} + \frac{\cos^2\gamma}{\frac{v^2}{c^2}\epsilon_3 - 1} + 1 = 0.$$

Clearing away the fractions and rejecting the factor v^2, we are left with a quadratic equation in v^2.

It should be noticed that $\mathbf{D}$ and $\mathbf{B}$ are perpendicular to the direction of propagation, but $\mathbf{E}$ is not.

Note 2, § 7

The solution represented by (17) and (20) can be verified by direct substitution. In order to obtain it one may proceed as follows:

The solenoidal distribution of the electric force $\mathbf{E}$ and the magnetic force $\mathbf{H}$ will be insured if we put

$$\mathbf{E} = \text{curl}\ \mathbf{E}', \quad \mathbf{H} = \text{curl}\ \mathbf{H}';$$

indeed, this implies $\quad \text{div}\ \mathbf{E} = \text{div curl}\ \mathbf{E}' = 0,$

$$\text{div}\ \mathbf{H} = \text{div curl}\ \mathbf{H}' = 0.$$

The vectors $\mathbf{E}'$ and $\mathbf{H}'$ have to be properly chosen in any particular case. They are of the kind called vector potentials.

Now, by putting $E'_z = - H'_y = \frac{c}{n} a \sin n\left(t - \frac{x}{c}\right)$,

$$E'_x = E'_y = H'_x = H'_z = 0,$$

where a is a slowly variable function of y and z, as specified in the text, we obtain the values (17) and (20). That Maxwell's equations are satisfied by these, to the degree of approximation here required, is found by remarking that

$$\text{curl } \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \text{curl } \mathbf{L}, \quad \text{curl } \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = \text{curl } \mathbf{N},$$

where
$$\mathbf{L} = \text{curl } \mathbf{H}' - \frac{1}{c}\frac{\partial \mathbf{E}'}{\partial t}, \quad \mathbf{N} = \text{curl } \mathbf{E}' + \frac{1}{c}\frac{\partial \mathbf{H}'}{\partial t}.$$

From these latter equations one finds, for the components of **L** and **N**,

$$L_x = \frac{c}{n}\frac{\partial a}{\partial z} \sin n\left(t - \frac{x}{c}\right), \quad L_y = 0, \quad L_z = 0,$$

$$N_x = \frac{c}{n}\frac{\partial a}{\partial y} \sin n\left(t - \frac{x}{c}\right), \quad N_y = 0, \quad N_z = 0,$$

showing that the components of curl **L** and curl **N** are either zero or negligible on account of a factor which is a small quantity of the second order.

Just as we have in (17) and (20) the representation of waves of variable amplitude with **E** in the direction of y and **H** in the direction of z, similar formulæ may be written down in which these directions are interchanged, and by combining the solution thus obtained (with a variable amplitude b) with (17) and (20), a solution is obtained in which both the amplitude and the polarization vary over the wave-front.

The method which we used here may also be applied to the propagation of curved waves that are laterally limited. Let S be any curved surface and let the position of a point Q on that surface be determined by two coördinates v and w chosen in such a way that the lines $v = \text{const.}$ and $w = \text{const.}$ are the lines of curvature. Let us draw the normal QN to a definite side of S, and let u be the length of a distance QP along that normal; then u, v, w may be taken for the coördinates of P.

At the point P we may distinguish three directions which we may call the directions u, v, w; namely, the directions in which P moves when one of the coördinates changes in the positive sense, the other two remaining constant. These directions, the first of which coincides with PN, are at right angles to each other, and v, w and QN may be chosen in such a way that the directions u, v, w may be made to coincide with those of x, y, z by a rotation.

The lengths of the displacements of P which correspond to positive increments du, dv, dw may be represented by $\alpha\,du$, $\beta\,dv$, $\gamma\,dw$, where α, β, γ are positive quantities that are functions of the coördinates. It is clear that $\alpha = 1$, but we shall retain this factor for the sake of symmetry.

As to β and γ, they will be pure numbers if the coördinates v and w have the dimensions of a length, and we shall suppose v and w to be so chosen that, throughout the space which we want to consider, β and γ shall be comparable to unity, being neither very large nor very small numbers. The equations of the electromagnetic field in the coördinates u, v, w are easily found when we remember that the two groups of Maxwell's equations are contained in the general formulæ

$$\int H_s ds = \frac{1}{c}\int \dot{D}_n d\sigma$$

and

$$\int E_s ds = -\frac{1}{c}\int \dot{B}_n d\sigma,$$

or, for the ether,

$$\int H_s ds = \frac{1}{c}\int \dot{E}_n d\sigma, \tag{291}$$

$$\int E_s ds = -\frac{1}{c}\int \dot{H}_n d\sigma. \tag{292}$$

In these equations the integrals on the left-hand side are to be taken along a closed line s, and the integrals on the right-hand side over a surface σ having s for its boundary, the normal n to σ being drawn in such a direction that it corresponds to the positive direction along s.

Applying (291) and (292) to infinitely small quadrangles, each of which has its sides in the directions of two of the coördinates, and denoting by E_u, E_v, E_w, H_u, H_v, H_w the components of $\mathbf{E}$ and $\mathbf{H}$ in the directions of the coördinates, we find that

$$\frac{\partial(\gamma H_w)}{\partial v} - \frac{\partial(\beta H_v)}{\partial w} = \frac{\beta\gamma}{c}\frac{\partial E_u}{\partial t}, \tag{293}$$

$$\frac{\partial(\alpha H_u)}{\partial w} - \frac{\partial(\gamma H_w)}{\partial u} = \frac{\gamma\alpha}{c}\frac{\partial E_v}{\partial t}, \tag{294}$$

$$\frac{\partial(\beta H_v)}{\partial u} - \frac{\partial(\alpha H_u)}{\partial v} = \frac{\alpha\beta}{c}\frac{\partial E_w}{\partial t}, \tag{295}$$

and that

$$\frac{\partial(\gamma E_w)}{\partial v} - \frac{\partial(\beta E_v)}{\partial w} = -\frac{\beta\gamma}{c}\frac{\partial H_u}{\partial t}, \tag{296}$$

$$\frac{\partial(\alpha E_u)}{\partial w} - \frac{\partial(\gamma E_w)}{\partial u} = -\frac{\gamma\alpha}{c}\frac{\partial H_v}{\partial t}, \tag{297}$$

$$\frac{\partial(\beta E_v)}{\partial u} - \frac{\partial(\alpha E_u)}{\partial v} = -\frac{\alpha\beta}{c}\frac{\partial H_w}{\partial t}. \tag{298}$$

We shall deduce from these equations that under proper circumstances there can exist a propagation of vibrations in which the surfaces $u = \text{const.}$ are the wave-fronts, the expressions for $E_u \cdots H_w$ all containing the factor

$$e^{in\left(t-\frac{u}{c}\right)}, \tag{299}$$

and the amplitudes by which these factors are multiplied being slowly varying functions of u, v, w. The condition for this state of things is that the radii of curvature of the wave-fronts are large in comparison with the wave-length λ; and when a quantity is said to be slowly varying, this means that it changes only by a small fraction of its magnitude when we pass over the distance of a wave-length. The quantities β and γ are of this kind, and if l is a length of the order of magnitude of the radii of curvature, or of the distance over which one must go forward in order to find a change in the amplitudes comparable with these themselves, a differentiation of β, γ, or one of the amplitudes with respect to u, v, w introduces a factor of the order $1/l$, whereas differentiation of (299) with respect to u introduces one of the order $1/\lambda$.

Among the terms of the above equations we can distinguish some that may be called principal terms, of the first order, and others of the second order that are of magnitude λ/l when compared with them. We shall neglect quantities of the third order, and in doing so we shall remember that all differentiations that are applied to a slowly varying quantity raise by one the order of magnitude of that quantity.

Now there are two solutions of the equations. In the first E_v and H_w are the only components that are of the first order,

$$E_v = a\,e^{in\left(t-\frac{u}{c}\right)}, \quad H_w = a'\,e^{in\left(t-\frac{u}{c}\right)}.$$

With the approximations just mentioned (putting now $\alpha = 1$) equations (294) and (298) become

$$-\frac{\partial(\gamma H_w)}{\partial u} = \frac{\gamma}{c}\frac{\partial E_v}{\partial t}, \quad \frac{\partial(\beta E_v)}{\partial u} = -\frac{\beta}{c}\frac{\partial H_w}{\partial t},$$

or, after substitution of the above values of E_v and H_w,

$$a = a' + i\frac{c}{n\gamma}\frac{\partial(\gamma a')}{\partial u}, \tag{300}$$

$$a' = a + i\frac{c}{n\beta}\frac{\partial(\beta a)}{\partial u}. \tag{301}$$

If, now, on the right-hand side of (300) we replace a' in the first term by the expression (301) and in the second simply by a, which amounts to the omission of a third-order term, we find

$$\frac{1}{\beta}\frac{\partial(\beta a)}{\partial u}+\frac{1}{\gamma}\frac{\partial(\gamma a)}{\partial u}=0,$$

or

$$2\frac{\partial \log a}{\partial u}+\frac{\partial \log(\beta\gamma)}{\partial u}=0,$$

$$a=\frac{p}{\sqrt{\beta\gamma}}, \tag{302}$$

where p is independent of u and may be any slowly varying function in the wave-front.

In reality $\frac{p}{\sqrt{\beta\gamma}}$ is only the first term of a series, the second term of which is of the order λ/l with respect to the first. We cannot deduce this second term from (300) and (301), because these equations are not exact, one of the terms having been omitted in both (294) and (298). But the expression (302) is sufficient for our purpose, because it shows us the law according to which, for very small values of λ/l, the amplitude varies in the direction of propagation when the waves are curved.

Similarly, and with the same restriction, we find, from (300) and (301),

$$a'=\frac{p}{\sqrt{\beta\gamma}},$$

so that, to a first approximation, the amplitudes of E_v and H_w are equal, as they are in the case of plane waves.

Returning now to equations (293)–(298), we may remark that the terms

$$\frac{\partial(\alpha H_u)}{\partial v} \quad \text{and} \quad \frac{\partial(\alpha E_u)}{\partial w},$$

in (295) and (297), may be omitted as being of the third order. We can therefore satisfy these equations by putting

$$E_w=0, \quad H_v=0,$$

whereas E_u and H_u are found from (293) and (296):

$$E_u=-i\frac{c}{\beta\gamma n}\frac{\partial}{\partial v}\left(p\sqrt{\frac{\gamma}{\beta}}\right)e^{in\left(t-\frac{u}{c}\right)},$$

$$H_u=-i\frac{c}{\beta\gamma n}\frac{\partial}{\partial w}\left(p\sqrt{\frac{\beta}{\gamma}}\right)e^{in\left(t-\frac{u}{c}\right)}.$$

We find similar formulæ for the second solution, in which E_w and H_v are the only components that are of the first order. It need scarcely be added that the final formulæ for the electromagnetic field are obtained by taking the real parts of the complex expressions for the components of E and H.

As to the variation of the amplitudes in the direction of propagation, we may remark that if R_1 and R_2 are the radii of curvature of the wave-front S from which we started, β and γ will be proportional to $R_1 + u$ and $R_2 + v$; that is, to the radii of curvature r_1 and r_2 of the wave-front passing through the point considered. Thus the amplitudes vary as $\dfrac{1}{\sqrt{r_1 r_2}}$ and the flow of energy as $\dfrac{1}{r_1 r_2}\cdot$ In the case of spherical waves this becomes $1/r^2$ for $r_1 = r_2 = r$, and it is easily found that in the case of cylindrical waves the flow of energy is inversely proportional to the first power of the one radius of curvature.

Note 3, § 10

In working out the theory of the method of the rotating mirror we shall suppose the angular velocity ω to be so small that all terms containing its square may be neglected. We shall also introduce some further simplifications.

Let the mirror be placed at the origin O of coördinates, its surface passing through OZ, which we take as the axis of rotation, and through the line which, in the plane XOY, makes an angle θ with OX. We shall suppose that, at $t = 0$, θ has the value $\frac{1}{4}\pi$ and that during the phenomena which we have to consider, t is so small that the term ωt in

$$\theta = \tfrac{1}{4}\pi + \omega t \tag{303}$$

can be treated as infinitely small.

Let the illuminated slit be placed at the focus of a collimating lens, so that the mirror receives a beam of parallel rays having the direction of OY. Then, if the mirror is held fixed in the position $\theta = \frac{1}{4}\pi$, the reflected rays have the direction OX. After having passed through a ponderable dispersive medium extending from $x = p$ to $x = q$, they are thrown back by a fixed mirror placed normally to OX.

In the experiments an image of the slit is formed on or near the surface of this mirror. This may be brought about by a slight displacement of the slit, so that the rays falling on the revolving mirror are not exactly parallel. We shall, however, avoid this complication, and we shall rather suppose that the parallel rays, after having passed

through the dispersive medium, fall on a convex lens L placed before the plane fixed mirror M, the focus of L lying in the surface of M. This combination, L, M, may be called the "inverting" arrangement.

The rotation of the first mirror will produce certain small changes in the beams of light with which we are concerned, and for the solution of the problem we want suitable expressions for the electric force **E** and the magnetic force **H** existing in them. For the sake of simplicity we shall suppose the electric force to have the direction of OZ; that of the magnetic force will then be in the plane XOY.

We shall take into account the dispersive property of the ponderable medium by assuming that, in the relation $\mathbf{D} = \epsilon \mathbf{E}$, the dielectric constant ϵ is a function of the frequency n. Then the wave-velocity

$$v = \frac{c}{\sqrt{\epsilon}} \tag{304}$$

will also depend on n, and according to what has been said in § 8 the group-velocity w will be determined by the relation

$$\frac{1}{w} = \frac{1}{v} + \frac{nv}{2\,c^2}\frac{d\epsilon}{dn}. \tag{305}$$

Now suitable expressions for the beam which, after the first reflection, is propagated in the ponderable medium are found to be

$$\left.\begin{aligned} E_z &= a\,e^{in\left(t-\frac{x}{v}\right)+i\kappa y\left(t-\frac{x}{w}\right)-\alpha v x+i\beta y-i\alpha n y^2} \\ H_x &= \left[-i\kappa\frac{c}{n^2}-\kappa\frac{c}{n}\left(t-\frac{x}{w}\right)-\beta\frac{c}{n}+2\,\alpha c y\right]E_z \\ H_y &= \left[-\frac{c}{v}+i\alpha\frac{cv}{n}+\kappa\frac{c}{n}\left(\frac{1}{v}-\frac{1}{w}\right)y\right]E_z \end{aligned}\right\}, \tag{306}$$

where κ, α, β are constants proportional to ω. Indeed, if we neglect quantities of the order ω^2, these values can be shown by direct substitution to satisfy the equations

$$\operatorname{curl}\mathbf{E} = -\frac{1}{c}\dot{\mathbf{H}}, \quad \operatorname{curl}\mathbf{H} = \frac{1}{c}\dot{\mathbf{D}} = \frac{1}{c}\epsilon\dot{\mathbf{E}};$$

only, since the frequency in (306) is $n + \kappa y$, one must replace ϵ by

$$\epsilon + \kappa y\frac{d\epsilon}{dn},$$

for which, by virtue of (304) and (305), we may write

$$\frac{c^2}{v^2} + \frac{2\,c^2}{nv}\left(\frac{1}{w} - \frac{1}{v}\right)\kappa y.$$

Corresponding to (306) we have the following expressions for the reflected beam before it reaches the ponderable medium:

$$\left.\begin{aligned} E_z &= a\, e^{in\left(t-\frac{x}{c}\right)+i\kappa y\left(t-\frac{x}{c}\right)-\alpha cx+i\beta y-i\alpha ny^2} \\ H_x &= \left[-i\kappa\frac{c}{n^2}-\kappa\frac{c}{n}\left(t-\frac{x}{c}\right)-\beta\frac{c}{n}+2\,\alpha cy\right]E_z \\ H_y &= \left(-1+i\alpha\frac{c^2}{n}\right)E_z \end{aligned}\right\}. \qquad (307)$$

We shall use expressions much like (306) and (307) for the other beams of light, and we shall have to determine for each of them the constants κ, α, β. The final result will be seen to depend mainly on the values of β. As to the amplitudes a, in general complex quantities, these are of no importance, for it is easily seen that a constant factor in the amplitude or a constant term in the phase cannot have any influence on the course of the waves.

Let the incident light (I) be given by

$$E_z = a\, e^{in\left(t-\frac{y}{c}\right)}, \quad H_x = E_z, \quad H_y = 0, \qquad (308)$$

and let us try to determine the amplitude a_r and the constants κ, α, β in the expressions (307) for the reflected light (II).

If we suppose the revolving mirror to consist of a perfectly conducting material, which implies that there can be no electromagnetic field in its interior, the condition at the surface, the only one that we have to consider, is that the normal component of the magnetic induction, that is, of **H**, must vanish. Hence the quantity

$$H_y \cos\theta - H_x \sin\theta,$$

and therefore also the expression

$$H_y - H_x \tan\theta,$$

or, by virtue of (303), $\quad H_y - (1+2\,\omega t)H_x, \qquad (309)$

must be zero.

At the surface we have

$$y = x\tan\theta = (1+2\,\omega t)x.$$

Hence, if this is substituted in (308) and (307), and if the quantity (309) is calculated for the incident and the reflected beam, the sum of the results must vanish for all values of t and x.

Putting
$$e^{in\left(t-\frac{x}{c}\right)} = \sigma,$$
we find, for the incident beam,
$$E_z = \sigma a\, e^{-\frac{2\,in\omega}{c}xt} = \sigma a\left(1 - \frac{2\,in\omega}{c}\,xt\right),$$
$$H_y - (1 + 2\,\omega t)H_x = \sigma a\left(-1 - 2\,\omega t + \frac{2\,in\omega}{c}\,xt\right),$$
and for the reflected beam,
$$E_z = \sigma a_r\, e^{i\kappa x\left(t-\frac{x}{c}\right) - \alpha c x + i\beta x - i\alpha n x^2}$$
$$= \sigma a_r\left[1 + (i\beta - \alpha c)x + i\kappa xt - i\left(\frac{\kappa}{c} + \alpha n\right)x^2\right],$$
$$H_y - (1 + 2\,\omega t)H_x = \left[-1 + i\alpha\frac{c^2}{n} + i\kappa\frac{c}{n^2} + \beta\frac{c}{n}\right.$$
$$\left.+ \kappa\frac{c}{n}t - \left(\frac{\kappa}{n} + 2\,\alpha c\right)x\right]E_z$$
$$= \sigma a_r\left[-1 + i\alpha\frac{c^2}{n} + i\kappa\frac{c}{n^2} + \beta\frac{c}{n} + \kappa\frac{c}{n}t\right.$$
$$\left.-\left(i\beta + \frac{\kappa}{n} + \alpha c\right)x - i\kappa xt + i\left(\frac{\kappa}{c} + \alpha n\right)x^2\right].$$

The surface-condition will be satisfied if
$$a_r = -a,$$
$$\kappa = -\frac{2\,n}{c}\,\omega, \quad \alpha = \frac{2}{c^2}\,\omega,$$
$$\beta_{II} = 0.$$

Beam II, falling on the first boundary, $x = p$, of the ponderable medium, will give rise to a transmitted beam, to be represented by (306) and to a reflected one, the expressions for which are obtained if in (307) we replace x by $-x$ and H_y by $-H_y$; thus, omitting the formula for H_x,
$$E_z = a\, e^{in\left(t+\frac{x}{c}\right) + i\kappa y\left(t+\frac{x}{c}\right) + \alpha c x + i\beta y - i\alpha n y^2},$$
$$H_y = \left(1 - i\alpha\frac{c^2}{n}\right)E_z.$$

So far as it is necessary we shall distinguish the quantities belonging to the three beams by the suffixes i (incident), r (reflected), and t (transmitted), and we have now to apply the conditions that at

the surface of separation E_z and H_y must be continuous (the continuity of H_x being implied in these). It is easily seen that this requires that in the exponents occurring in the expressions for E_z the coefficients of t, yt, and y^2 must be equal, so that n, κ, and α must be the same in the three cases. Furthermore, if we put

$$e^{int+i\kappa yt-i\alpha ny^2}=\sigma,$$

$$a_i e^{-\frac{inp}{c}}=b_i,\quad a_r e^{\frac{inp}{c}}=b_r,\quad a_t e^{-\frac{inp}{v}}=b_t,$$

we have

$$E_{zi}=\sigma b_i e^{-i\kappa\frac{p}{c}y-\alpha cp+i\beta_i y}$$

$$=\sigma b_i\Big(1-\alpha cp-i\kappa\frac{p}{c}y+i\beta_i y\Big),$$

$$H_{yi}=\sigma b_i\Big(-1+i\alpha\frac{c^2}{n}+\alpha cp+i\kappa\frac{p}{c}y-i\beta_i y\Big),$$

$$E_{zr}=\sigma b_r\Big(1+\alpha cp+i\kappa\frac{p}{c}y+i\beta_r y\Big),$$

$$H_{yr}=\sigma b_r\Big(1-i\alpha\frac{c^2}{n}+\alpha cp+i\kappa\frac{p}{c}y+i\beta_r y\Big),$$

$$E_{zt}=\sigma b_t\Big(1-\alpha vp-i\kappa\frac{p}{w}y+i\beta_t y\Big),$$

$$H_{yt}=\sigma b_t\Big[-\frac{c}{v}+i\alpha\frac{cv}{n}+\alpha cp+\kappa\frac{c}{n}\Big(\frac{1}{v}-\frac{1}{w}\Big)y+i\kappa\frac{cp}{vw}y-i\beta_t\frac{c}{v}y\Big].$$

Applying the boundary conditions to the case $\omega=0$, $\kappa=0$, $\alpha=0$, $\beta=0$, we find

$$b_i:b_r:b_t=(v+c):(v-c):2\,v,$$

and these ratios, though in the real case they differ from the true ones by quantities of the order ω, may be used in terms containing κ and β.

If now we consider only those terms in the equations

$$E_{zi}+E_{zr}=E_{zt},\quad H_{yi}+H_{y}\ =H_{yt}$$

which contain the factor y, and which, taken by themselves, must satisfy the equations, we obtain two relations for the determination of β_i and β_t. For the second of these quantities the result is

$$\beta_{III}=\beta_{II}+\kappa p\left(\frac{1}{w}-\frac{1}{c}\right),\tag{310}$$

where a term $-i\kappa\dfrac{c(w-v)}{nw(v+c)}$ has been neglected because its ratio to the term $\kappa p\left(\dfrac{1}{w}-\dfrac{1}{c}\right)$ is of the order of magnitude $\dfrac{\lambda}{p}$, if λ is the wave-length.

We can examine in exactly the same way what takes place at the surface $x = q$, where the light leaves the ponderable medium. The incident light is now to be represented by expressions of the form (306), the transmitted light (IV) by (307), and the reflected beam by

$$E_z = a\, e^{in\left(t+\frac{x}{v}\right)+i\kappa y\left(t+\frac{x}{w}\right)+\alpha r x - i\beta y - i\alpha n y^2},$$

$$H_y = \left[\frac{c}{v} - i\alpha\frac{cv}{n} - \kappa\frac{c}{n}\left(\frac{1}{v}-\frac{1}{w}\right)y\right]E_z.$$

Again, the values of n, κ, and α are the same for the three beams, and we find

$$\beta_{IV} = \beta_{III} + \kappa q\left(\frac{1}{c}-\frac{1}{w}\right)$$

and, combining this with (310),

$$\beta_{IV} = \beta_{II} + \kappa l\left(\frac{1}{c}-\frac{1}{w}\right),$$

if $l = q - p$ is the length of the column of ponderable matter.

It will be convenient, before considering the inversion of beam (IV), to shift the origin of coördinates to a point just in front of the lens. We have then to replace x by $x + L$, if L is the distance from the revolving mirror to the lens. After this change we may still use (307), if only we replace β by $\beta - \frac{\kappa L}{c}$. By this the constants for beam (IV) become

$$\kappa = -\frac{2n}{c}\,\omega, \quad \alpha = \frac{2}{c^2}\,\omega, \quad \beta'_{IV} = \beta_{II} - \kappa L\cdot\frac{1}{c} + \kappa l\left(\frac{1}{c}-\frac{1}{w}\right). \tag{311}$$

It would be difficult exactly to analyze what goes on in the inverting arrangement. Nevertheless expressions for the inverted beam (V) which probably are sufficiently near the truth are easily found.

Let OX' and OY' be new axes in directions opposite to those of OX and OY. Then (V) is represented by formulæ of the form (307); namely,

$$\left.\begin{aligned} E_z &= a\, e^{in\left(t-\frac{x'}{c}\right)+i\kappa y'\left(t-\frac{x'}{c}\right)-\alpha c x'+i\beta y'-i\alpha n y'^2} \\ H_{x'} &= \left[-i\kappa\frac{c}{n^2} - \kappa\frac{c}{n}\left(t-\frac{x'}{c}\right) - \beta\frac{c}{n} + 2\,\alpha c y'\right]E_z \\ H_{y'} &= \left(-1 + i\alpha\frac{c^2}{n}\right)E_z \end{aligned}\right\}. \tag{312}$$

with the values (311) of κ, α, β that were found for beam (IV).

That these are the proper formulæ for the inverted beam may be seen by attending to the meaning of the terms with κ, α, β.

1. If $\kappa = 0$, $\alpha = 0$, the phase in (307) is the same in a plane passing through O and determined by

$$-n\frac{x}{c} + \beta y = 0. \tag{313}$$

Similarly, the wave-front in (312) is given by

$$-n\frac{x'}{c} + \beta y' = 0,$$

so that it coincides with the plane (313). This must be so because the inversion by the lens and the mirror placed at its focus leaves the direction of plane waves unchanged.

2. If $\kappa = 0$, $\beta = 0$, the equation of the wave-front is

$$-n\frac{x}{c} - \alpha n y^2 = 0$$

in one case, and

$$-n\frac{x'}{c} - \alpha n y'^2 = 0$$

in the other. The two are equally curved, but in opposite directions, as they are found to be by the rules of elementary optics.

3. The frequency is $n + \kappa y$ in (307) and $n + \kappa y'$ in (312). If in the first case it is greatest for positive values of y, it is greatest in the second case for positive values of y' corresponding to negative ones of y. This also could be expected, because a ray which first passes through the lens on one side of OX will, after the reflection, pass through it on the other side of that axis.

The problem of the propagation through the dispersive medium is the same for beam (V), represented by (312), as it was for beam (II), represented by (307). One can say at once that the formulæ (312) apply to the returning beam (VI) between the ponderable medium and the revolving mirror, if the constants are given the values (311), with the exception that β has to be increased by $\kappa l\left(\frac{1}{c} - \frac{1}{w}\right)\cdot$ If, finally, the origin is shifted back to its original position at the center of the rotating mirror, β is further altered by $-\kappa L\cdot\frac{1}{c}$, so that for the returning beam (VI) falling on the mirror we have (312) with the values

$$\kappa = -\frac{2n}{c}\omega,\quad \alpha = \frac{2}{c^2}\omega,\quad \beta = -2\kappa L\cdot\frac{1}{c} + 2\kappa l\left(\frac{1}{c} - \frac{1}{w}\right)\cdot \tag{314}$$

It remains to examine the last reflection. The reflected beam will be propagated nearly along OY'. It will differ more or less from

what it would be in the simple case $\omega = 0$, and presumably the modifications due to the rotation of the mirror will be similar to those which show themselves in the other beams of light. One is therefore led to try expressions like (312), replacing x' by y' and y' by $-x'$. On working out the problem we find that formulæ of this kind can satisfy the boundary condition, and that in the new beam $\kappa = 0$, $\alpha = 0$. For the sake of brevity we shall assume these values, writing for the reflected rays

$$E_z = a_r e^{in\left(t - \frac{y'}{c}\right) - i\beta_r x'}, \quad H_{y'} = -\beta_r \frac{c}{n} E_z, \quad H_{x'} = E_z. \tag{315}$$

We shall show that the condition at the surface can be satisfied if β_r is properly chosen.

The normal component of the magnetic force is

$$-H_{y'} \cos\theta + H_{x'} \sin\theta,$$

so that the quantity $\quad H_{y'} - (1 + 2\,\omega t)H_{x'}$

must vanish at the mirror, that is, for all values of t and x', if first we substitute

$$y' = x' \tan\theta = (1 + 2\,\omega t)x'.$$

Denoting by σ the factor $e^{in\left(t - \frac{x'}{c}\right)}$ and using the values (314) of κ and α, we find for beam (VI)

$$E_z = \sigma a\, e^{i\kappa x'\left(t - \frac{x'}{c}\right) - \alpha x' + i\beta x' - i\kappa n x'^2}$$

$$= \sigma a\left(1 - \frac{2}{c}\,\omega x' + i\beta x' - \frac{2\,in\omega}{c}\,x't\right),$$

$$H_{y'} - (1 + 2\,\omega t)H_{x'} = \left(-1 + \beta\frac{c}{n} - 2\,\omega t - \frac{2}{c}\,\omega x'\right)E_z$$

$$= \sigma a\left(-1 + \beta\frac{c}{n} - 2\,\omega t - i\beta x' + \frac{2\,in\omega}{c}\,x't\right),$$

and for the reflected light

$$E_z = \sigma a_r e^{-in\cdot\frac{2\,\omega x'}{c}t - i\beta_r x'} = \sigma a_r\left(1 - i\beta_r x' - \frac{2\,in\omega}{c}\,x't\right),$$

$$H_{y'} - (1 + 2\,\omega t)H_{x'} = \left(-1 - \beta_r\frac{c}{n} - 2\,\omega t\right)E_z$$

$$= \sigma a_r\left(-1 - \beta_r\frac{c}{n} - 2\,\omega t + i\beta_r x' + \frac{2\,in\omega}{c}\,x't\right).$$

These values agree with the surface-conditions if $a_r = -a$ and $\beta_r = -\beta$, or, on account of (314),

$$\beta_r = -\frac{4n\omega}{c}\left[L\frac{1}{c} - l\left(\frac{1}{c} - \frac{1}{w}\right)\right].$$

The second of equations (315) shows that the direction of the ultimate reflected beam deviates from the direction which it would have in the case of no rotation by an angle

$$\phi = -\frac{c}{n}\beta_r,$$

the positive direction of the deviation being from OX' toward OY', or from OX toward OY, and being therefore the same as the positive direction of the angular velocity ω. Thus there will be a deviation

$$\phi = 4\omega\left[L\frac{1}{c} - l\left(\frac{1}{c} - \frac{1}{\omega}\right)\right]$$

having the direction of the rotation of the mirror, as is really observed. Now, in discussing the experiment in the ordinary way, if this deviation ϕ has been measured, one infers from it that in the time during which the light has traveled to the distant mirror and back again the movable mirror has turned by an angle

$$2\omega\left[L\frac{1}{c} - l\left(\frac{1}{c} - \frac{1}{\omega}\right)\right],$$

and that therefore the time required for the propagation from one mirror to the other is

$$L\frac{1}{c} - l\left(\frac{1}{c} - \frac{1}{w}\right).$$

As the path in the ether is $L - l$, the time required by it is $\frac{L - l}{c}$, so that l/w is left for the propagation in the column of ponderable matter. Thus the velocity deduced from the experiment is really the group-velocity w, although after these complicated calculations we cannot say that this result is at all obvious at first sight.

Of course we could have considered, as was done in the lecture, the intervals of time during which the outgoing rays fall on the fixed mirror and the returning rays on the revolving one, and in doing so we should have to attend to the group-velocity in the dispersive medium. But the above calculations show (and this is easily understood) that in determining the position of the final

image we are not concerned with these intervals of time; it depends only on the direction of the wave-fronts.

Some further remarks may be added:

1. If light is reflected by a mirror receding with the velocity u in the direction of the normal, the change in the frequency is, by Doppler's principle,

$$-\frac{2\,nu}{c}\cos\chi,$$

where χ is the angle of incidence. This agrees with the change $\kappa y = -\frac{2\,n\omega y}{c}$, because, at the distance s from O, the velocity of a point of the mirror is $u = \omega s$, and for the position in which the mirror bisects the angle XOY, $y = s\cos\chi$.

2. The quantity by which i is multiplied in the exponent in (306) may be called the "phase," and if it is put equal to a constant k, we find the equation of a wave-front containing t as a parameter. For $y = 0$ the equation becomes

$$n\left(t - \frac{x}{v}\right) = k, \tag{316}$$

showing that the point of intersection of a wave with OX proceeds with the velocity v. Furthermore, at this point the derivative dx/dy, taken for a constant t, is found to be

$$\frac{dx}{dy} = \frac{v}{n}\left[\kappa\left(t - \frac{x}{w}\right) + \beta\right],$$

or, using (316),

$$\frac{dx}{dy} = \frac{\kappa}{n}\left(1 - \frac{v}{w}\right)x + \text{const.}$$

The latter of these formulæ determines the change in the direction of a wave during its propagation in a dispersive medium.

3. The term with y^2 in the exponent shows that plane waves, when reflected by a revolving mirror, become slightly curved, a result that can be verified by means of Huygens's construction. The sense of the curvature depends on the direction of the rotation. Corresponding to the curvature of the waves there is the term $-\alpha vx$, the meaning of which is that in the case of curved waves the amplitude varies along the normal, increasing in the direction of their concave side.

Note 4, § 17

By means of equations (80) Mr. Bateman has rigorously calculated the energy radiated by an electron moving periodically in a closed orbit of any size. He has been led to the interesting result

that for a full period the radiation is equal to that which would correspond to a flux of energy

$$S' = 2\,\psi\,\frac{\partial \operatorname{grad}\psi}{\partial t}, \tag{317}$$

where

$$\psi = \sqrt{1 - \frac{[v^2]}{c^2}}\,\phi,$$

so that we could get rid of the radiation by supposing that, in addition to Poynting's flow,

$$S = c[E \cdot H],$$

there is another flow equal and opposite to S′.

The calculation is as follows: Let x, y, z be the coördinates of a point P fixed in space, t the time for which we want to determine the field, t_e the time at which the electron reaches its effective position Q, x_e, y_e, z_e the coördinates in that position, $r = c(t - t_e)$ the distance QP, v the velocity of the electron at time t_e, and v_r its component in the direction QP. When the motion of the electron is given, t_e and consequently x_e, y_e, z_e, r, and v will be definite functions of x, y, z, t.

Differentiating the equation

$$(x - x_e)^2 + (y - y_e)^2 + (z - z_e)^2 - c^2(t - t_e)^2 = 0$$

with respect to t, we find

$$-[(x - x_e)v_x + (y - y_e)v_y + (z - z_e)v_z]\,\frac{\partial t_e}{\partial t} - c^2(t - t_e)\left(1 - \frac{\partial t_e}{\partial t}\right) = 0,$$

giving

$$\frac{\partial t_e}{\partial t} = \left(1 - \frac{v_r}{c}\right)^{-1}, \tag{318}$$

and, similarly,

$$\frac{\partial t_e}{\partial x} = -\,\frac{x - x_e}{cr}\left(1 - \frac{v_r}{c}\right)^{-1} = -\,\frac{x - x_e}{cr}\,\frac{\partial t_e}{\partial t}, \text{ etc.}$$

By this the derivatives of x_e, y_e, z_e, r, and v, all of which depend on t_e, likewise become known.

Now the electromagnetic field is determined by the derivatives of ϕ and **A**. We shall determine the flow of energy across a very large sphere, and it will suffice to retain in **E** and **H** only terms of order $1/r$. This means that we have not to differentiate the factor $1/r$.

Omitting it and the constant factor $e/4\,\pi$, we shall put

$$\phi = (1 - w_r)^{-1}, \quad \mathbf{A} = \mathbf{w}(1 - w_r)^{-1},$$

where for further abbreviation we have replaced v/c by w.

Denoting, on the right-hand side of the equations, by x, y, z the relative coördinates of P with respect to Q, that is, the quantities that were first called $x - x_e$, $y - y_e$, $z - z_e$, we have, for each of the quantities that have to be differentiated,

$$\frac{\partial}{\partial x} = -\frac{x}{cr}\frac{\partial}{\partial t}, \quad \frac{\partial}{\partial y} = -\frac{y}{cr}\frac{\partial}{\partial t}, \quad \frac{\partial}{\partial z} = -\frac{z}{cr}\frac{\partial}{\partial t}.$$

Differentiations with respect to t will be represented by dots. If now, in the expressions for the components of **E** and **H** we omit the factor $1/cr$, they take the form

$$E_x = x\dot{\phi} - r\dot{A}_x \text{ etc.,} \quad H_x = -y\dot{A}_z + z\dot{A}_y \text{ etc.,}$$

and from these we find, for the components of Poynting's flow, after some transformation, omitting the factor c,

$$S_x = E_yH_z - E_zH_y = \dot{\phi}[r^2\dot{A}_x - x(x\dot{A}_x + y\dot{A}_y + z\dot{A}_z)]$$
$$+ r[x(\dot{A}_x^2 + \dot{A}_y^2 + \dot{A}_z^2) - \dot{A}_x(x\dot{A}_x + y\dot{A}_y + z\dot{A}_z)], \text{ etc.} \quad (319)$$

As we have omitted, in both **E** and **H**, the factors $e/4\pi r$ and $1/cr$, and after that in **S** the factor c, the components of **S** will finally have to be multiplied by

$$\frac{e^2}{16\pi^2cr^4}.$$

We shall now integrate the values of S_x etc. over a full period T, beginning at some definite instant t_0. If t_{e0} is the effective time corresponding to this, that which corresponds to $t_0 + T$ will be $t_{e0} + T$, and if we take t_e as the integration variable, attending to the relation (318), we have for the integrals in question

$$\int_{t_{e0}}^{t_{e0}+T} (1 - w_r)S_x dt_e, \text{ etc.} \quad (320)$$

In the above value of S_x we must substitute $\phi = (1 - w_r)^{-1}$, $\dot{\phi} = \dot{w}_r(1 - w_r)^{-2}$, and similar expressions for $\dot{A}_x$ etc. Furthermore,

$$w_r = \frac{x}{r}w_x + \frac{y}{r}w_y + \frac{z}{r}w_z.$$

As we may consider the distances to the point P as infinitely great with respect to the dimensions of the orbit, we are justified, for the same reasons for which we did not differentiate $1/r$, in considering as constants x, y, z, and r (though they vary over ranges determined by the dimensions of the orbit), so long as P is kept fixed, taking for them the values that correspond to the position Q_0 of the electron at the time t_{e0}. This enables us to calculate in a simple manner

the energy passing through a great sphere having its center at Q_0. In this problem we are concerned with

$$S_r = \frac{x}{r} S_x + \frac{y}{r} S_y + \frac{z}{r} S_z,$$

or, by virtue of (319),

$$S_r = r^2(\dot{A}_x^2 + \dot{A}_y^2 + \dot{A}_z^2) - (x\dot{A}_x + y\dot{A}_y + z\dot{A}_z)^2,$$

and we have to calculate

$$I = \int_{t_{e0}}^{t_{e0}+T} (1 - w_r) S_r dt_e, \tag{321}$$

corresponding to (320).

We shall now substitute in S_r the values

$$A_x = w_x(1 - w_r)^{-1}, \text{ etc.}$$

expressing at the same time the derivatives with respect to t in those with respect to t_e, which we shall denote by primes. Here the relation (318) must again be taken into account. Thus

$$\dot{A}_x = (1 - w_r)^{-1}A'_x = (1 - w_r)^{-2}w'_x + (1 - w_r)^{-3}w_x w'_r, \text{ etc.},$$

$$\dot{A}_x^2 + \dot{A}_y^2 + \dot{A}_z^2 = (1 - w_r)^{-4}w'^2 + 2(1 - w_r)^{-5}(w \cdot w')w'_r$$
$$+ (1 - w_r)^{-6}w^2 w'^2_r,$$

$$x\dot{A}_x + y\dot{A}_y + z\dot{A}_z = r\,[(1 - w_r)^{-2}w'_r + (1 - w_r)^{-3}w_r w'_r]$$
$$= r(1 - w_r)^{-3}w'_r,$$

$$(1 - w_r)S_r = r^2[(1 - w_r)^{-3}w'^2 + 2(1 - w_r)^{-4}(w \cdot w')w'_r$$
$$- (1 - w_r)^{-5}(1 - w^2)w'^2_r]. \tag{322}$$

In order to find the total radiation we must integrate the expression (321) over the sphere described about O with radius r. We may, however, invert the order of the two operations, first integrating (322) over the sphere, for definite values of t_e and dt_e and then over a period. It is to be noted that in the first step of this calculation we do not take together amounts of energy that pass through different parts of the sphere simultaneously, for when t_e has a definite value, t will not be constant all over the surface.

At the time t_e which we choose the velocity w will have a definite direction. Let us take the axis of x in that direction, and let us also introduce polar coördinates, θ, χ, such that the coördinates of a point of the spherical surface are

$$x = r\cos\theta, \quad y = r\sin\theta\cos\chi, \quad z = r\sin\theta\sin\chi.$$

We then have

$$w_r = w\cos\theta, \quad w'_r = w'_x\cos\theta + w'_y\sin\theta\cos\chi + w'_z\sin\theta\sin\chi.$$

After having substituted these values in (322) we must multiply by an element of the surface $d\sigma = r^2 \sin\theta\, d\theta\, d\chi$ and then integrate with respect to χ between 0 and 2π, and with respect to θ between 0 and π. The first integration is easily performed and we find

$$2\pi r^4 \int \{(1-w_r)^{-3} w'^2 + 2(1-w_r)^{-4}(w\cdot w')w'_x \cos\theta$$
$$-(1-w_r)^{-5}(1-w^2)[w'^2_x \cos^2\theta + \tfrac{1}{2}(w'^2_y + w'^2_z)\sin^2\theta]\} \sin\theta\, d\theta,$$

or, if we put $\cos\theta = -u$,

$$2\pi r^4 \int_{-1}^{+1} \{(1+wu)^{-3} w'^2 - 2(1+wu)^{-4}(w\cdot w')\, w'_x u$$
$$-(1+wu)^{-5}(1-w^2)[\tfrac{1}{2}(w'^2 - w'^2_x) + \tfrac{1}{2}(3\, w'^2_x - w'^2)u^2]\}\, du. \quad (323)$$

The integrals of $(1+wu)^{-3}$, $(1+wu)^{-4}u$, $(1+wu)^{-5}$, and $(1+wu)^{-5}u^2$ have the values

$$\frac{2}{(1-w^2)^2},\quad -\frac{8\,w}{3(1-w^2)^3},\quad \frac{2(1+w^2)}{(1-w^2)^4},\quad \text{and}\quad \frac{2+10\,w^2}{3(1-w^2)^4},$$

and (323) becomes

$$\int (1-w_r) S_r\, d\sigma = \frac{8}{3}\pi r^4 \left[\frac{w'^2}{(1-w^2)^2} + \frac{(w\cdot w')^2}{(1-w^2)^3}\right]. \quad (324)$$

Since w had the direction x, the product ww'_x has been represented by $(w\cdot w')$, and the result now has a form that is independent of the temporary choice of the axes.

Integrating (324) with respect to t_e, and restoring the factor that has been omitted, we find, for the radiation during a period, Liénard's formula

$$\frac{ce^2}{6\pi}\int \left[\frac{v'^2}{(c^2-v^2)^2} + \frac{(v\cdot v')^2}{(c^2-v^2)^3}\right] dt.$$

We have substituted $w = \frac{v}{c}$, $w' = \frac{v'}{c}$, so that v' is the acceleration, and we have written t instead of t_e. Indeed, in calculating the integral we need no longer think of what was meant by the suffix e.

For slow motions we are led back to the formula that was given in the lecture.

The radiation corresponding to the flux of energy $\mathbf{S}'$ mentioned in the beginning of this note can be calculated along exactly the same lines.

To begin with we shall replace (317) by

$$\mathbf{S}' = -2\frac{\partial\psi}{\partial t}\operatorname{grad}\psi;$$

this amounts to the omission of the flux $2\frac{\partial}{\partial t}(\psi \operatorname{grad} \psi)$, the integral of which over a full period is zero. If now, replacing ψ by $(1-w^2)^{\frac{1}{2}}(1-w_r)^{-1}$, and $\frac{\partial}{\partial x}$ by $-xr\frac{\partial}{\partial t}$, etc., we write, for the components of the new $\mathbf{S}'$,

$$S'_x = 2\,xr\left\{\frac{\partial}{\partial t}[(1-w^2)^{\frac{1}{2}}(1-w_r)^{-1}]\right\}^2, \text{ etc.,}$$

we have omitted in $\mathbf{S}'$ the same factor as was left aside in $\mathbf{S}$.

We have next, when we introduce again the derivatives with respect to t_e,

$$(1-w_r)S'_r = 2\,r^2(1-w_r)^{-1}[-(1-w^2)^{-\frac{1}{2}}(1-w_r)^{-1}(w\cdot w') \\ +(1-w^2)^{\frac{1}{2}}(1-w_r)^{-2}w'_r]^2,$$

and if this is worked out and compared with (322),

$$(1-w_r)S'_r = -2(1-w_r)S_r \\ +2\,r^2(1-w_r)^{-3}[w'^2+(1-w^2)^{-1}(w\cdot w')^2].$$

Now integrate over the sphere. The last term gives thrice the expression on the right-hand side of (324); thus,

$$\int(1-w_r)S'_r d\sigma = \int(1-w_r)S_r d\sigma,$$

by which the theorem is proved.

Of course, if we change the flow of energy by the addition of the flux $\mathbf{S}'$, we must expect that the other components of the stress-energy tensor will have to be modified likewise. Mr. Bateman has examined this question and some further consequences of the theory in his paper, "The Stress-Energy Tensor in Electromagnetic Theory and a New Law of Force," *Phys. Review*, Vol. 20 (1922), p. 243.

Note 5, § 26

It is well known that the interference phenomena produced when X-rays fall on a crystal can be interpreted as due to the reflection by a series of equidistant crystallographic planes. Let us therefore consider one such plane V, supposing that the molecules which are arranged in it in a regular pattern are so close together that they may be replaced by a continuous distribution.

Let the rays come from the point P and let Q be the point for which we wish to determine the reflected vibrations. Moreover, we shall take as origin of two axes OX and OY in the plane the point O of regular reflection, placing OX along the intersection of the planes POQ and V. We shall draw OZ toward the side of the incident rays.

Let the angle of incidence POZ be θ, $PO = l$, $OQ = l'$, r and r' the distances of a point in the plane V from P and Q, $d\sigma$ an element of the plane, N the number of molecules per unit of area. The vibrations at Q in any chosen direction, in so far as they are due to the molecules lying in $d\sigma$, may be represented by an expression of the form

$$\frac{Na}{rr'} \cos n\left(t - \frac{r + r'}{c} + p\right) d\sigma. \tag{325}$$

The factor $1/rr'$ has been introduced here because the amplitude of the vibration which falls on a particle, and therefore also that of the motion excited in it, is inversely proportional to r, and because, similarly, in the propagation toward Q the amplitude changes in the inverse ratio of r'.

We have now to integrate the above expression, in which a also may change from point to point, over the part of the plane that receives the rays.

The sum $r + r'$, which we shall call s, is a minimum $s_0 = l + l'$ at the point O, and the curves $s =$ const. are ellipses surrounding this point. We suppose the boundary of the illuminated area not to coincide for a finite length with one of these ellipses; then the greatest value s_m will be reached at a single point of the boundary.

Following the same method that was used in the lecture, we put

$$\int_{s_0}^{s} \frac{Na}{rr'} d\sigma = F(s), \tag{326}$$

the integral being extended to the part of the area lying within an ellipse s. The integral

$$\int_{s}^{s+ds} \frac{Na}{rr'} d\sigma,$$

taken over the part between two consecutive ellipses, may be represented by

$$\frac{dF(s)}{ds} ds,$$

and for the integral which we have to determine we may write

$$I = \int_{s_0}^{s_m} \frac{dF(s)}{ds} \cos n\left(t - \frac{s}{c} + p\right) ds,$$

or, after an integration by parts,

$$I = -\frac{c}{n}\left|\frac{dF(s)}{ds} \sin n\left(t - \frac{s}{c} + p\right)\right|_{s=s_0}^{s=s_m} + \frac{c}{n}\int_{s_0}^{s_m} \frac{d^2F(s)}{ds^2} \sin n\left(t - \frac{s}{c} + p\right) ds. \tag{327}$$

We may here confine ourselves to the first term just as we did in the discussion of the expression (101) on page 73. It can be shown that, since the value s_m is found at a single point of the boundary, the value of $\frac{dF(s)}{ds}$ for that value vanishes. As to the value at O, it can be calculated if one remarks that for values of s very near s_0 we have full ellipses within the illuminated area, the point O being supposed to be at some distance from the boundary.

The integral over one of these small ellipses is easily found. As the coördinates of P are $-l \sin\theta,\ 0,\ l\cos\theta$, and those of Q are $l'\sin\theta,\ 0,\ l'\cos\theta$, we have

$$r^2 = l^2 + 2\,xl\sin\theta + x^2 + y^2,$$

and for small values of x and y

$$r = l + x\sin\theta + \frac{1}{2\,l}(x^2\cos^2\theta + y^2).$$

Similarly $$r' = l' - x\sin\theta + \frac{1}{2\,l'}(x^2\cos^2\theta + y^2).$$

From this we find

$$\frac{1}{2}\left(\frac{1}{l} + \frac{1}{l'}\right)(x^2\cos^2\theta + y^2) = s - s_0,$$

so that the semi-axes of the ellipse corresponding to a small value of $s - s_0$ are

$$\frac{1}{\cos\theta}\sqrt{\frac{2\,ll'(s-s_0)}{l+l'}},\quad \sqrt{\frac{2\,ll'(s-s_0)}{l+l'}}.$$

It will suffice to retain in the integral (326) terms of order $s - s_0$ only, because we have to calculate the derivative of $F(s)$ for $s = s_0$. We may therefore consider Na/rr' as constant, multiplying its value at O by the area of the ellipse.

Thus $$F(s) = \frac{2\,\pi Na}{(l+l')\cos\theta}(s - s_0),$$

by which (327) becomes

$$I = \frac{2\,\pi Nac}{n(l+l')\cos\theta}\sin n\left(t - \frac{s}{c} + p\right).$$

If this is compared with (325), supposing $l' = l$, it is seen that, apart from a change in phase of a quarter period, the reflected vibration is what we should have if the molecules contained in an area

$$\frac{\lambda l}{2\cos\theta} \tag{328}$$

were concentrated at O.

The effective area (328) is of the order of magnitude of the zones into which the plane is divided by the ellipses $s - s_0 = \lambda,\ 2\,\lambda$, etc.

(Fresnel's zones). Under ordinary circumstances it is very small (for example, $4 \cdot 10^{-7}$ cm.^{2}, when $\lambda = 10^{-8}$ centimeters, $l = 40$ centimeters, and $\theta = 60°$), yet it contains a very great number of molecules.

Note 6, § 27*

The electromagnetic field that is produced, in the absence of all ponderable matter, by a given distribution of electromotive and magnetomotive forces $\mathbf{E}_e$ and $\mathbf{H}_e$ (which may be any functions of the time) can be determined as follows:

The fundamental equations are

$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \quad \operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t},$$

$$\mathbf{D} = \mathbf{E} + \mathbf{E}_e, \qquad \mathbf{B} = \mathbf{H} + \mathbf{H}_e.$$

From these we find

$$\operatorname{curl}\operatorname{curl} \mathbf{E} = -\frac{1}{c}\operatorname{curl} \dot{\mathbf{B}} = -\frac{1}{c}(\operatorname{curl} \dot{\mathbf{H}} + \operatorname{curl} \dot{\mathbf{H}}_e)$$

$$= -\frac{1}{c^2}\ddot{\mathbf{D}} - \frac{1}{c}\operatorname{curl} \dot{\mathbf{H}}_e = -\frac{1}{c^2}\ddot{\mathbf{E}} - \frac{1}{c^2}\ddot{\mathbf{E}}_e - \frac{1}{c}\operatorname{curl} \dot{\mathbf{H}}_e,$$

or, since we may write on the left-hand side (on account of $\operatorname{div} \mathbf{D} = 0$)

$$\operatorname{grad}\operatorname{div} \mathbf{E} - \Delta \mathbf{E} = -\operatorname{grad}\operatorname{div} \mathbf{E}_e - \Delta \mathbf{E},$$

$$\Delta \mathbf{E} - \frac{1}{c^2}\ddot{\mathbf{E}} = \frac{1}{c}\operatorname{curl} \dot{\mathbf{H}}_e - \operatorname{grad}\operatorname{div} \mathbf{E}_e + \frac{1}{c^2}\ddot{\mathbf{E}}_e. \tag{329}$$

Similarly

$$\Delta \mathbf{H} - \frac{1}{c^2}\ddot{\mathbf{H}} = -\frac{1}{c}\operatorname{curl} \dot{\mathbf{E}}_e - \operatorname{grad}\operatorname{div} \mathbf{H}_e + \frac{1}{c^2}\ddot{\mathbf{H}}_e. \tag{330}$$

If now two auxiliary vectors P and Q are determined by the equations

$$\Delta \mathbf{P} - \frac{1}{c^2}\ddot{\mathbf{P}} = \mathbf{E}_e, \quad \Delta \mathbf{Q} - \frac{1}{c^2}\ddot{\mathbf{Q}} = \mathbf{H}_e, \tag{331}$$

(329) and (330) are satisfied by

$$\mathbf{E} = \frac{1}{c}\operatorname{curl} \dot{\mathbf{Q}} - \operatorname{grad}\operatorname{div} \mathbf{P} + \frac{1}{c^2}\ddot{\mathbf{P}},$$

$$\mathbf{H} = -\frac{1}{c}\operatorname{curl} \dot{\mathbf{P}} - \operatorname{grad}\operatorname{div} \mathbf{Q} + \frac{1}{c^2}\ddot{\mathbf{Q}}.$$

* All formulæ in this note are vector equations. If ϕ is a scalar, grad ϕ is the vector whose components are $\frac{\partial \phi}{\partial x}$ etc. The operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

applied to a vector A, gives a new vector with the components ΔA_x etc. In the transformations use has been made of the relation

$$\operatorname{curl}\operatorname{curl} \mathbf{A} = \operatorname{grad}\operatorname{div} \mathbf{A} - \Delta \mathbf{A}.$$

A solution of (331) is

$$P = -\frac{1}{4\pi}\int\frac{[E_e]}{r}\,dS, \quad Q = -\frac{1}{4\pi}\int\frac{[H_e]}{r}\,dS,$$

where dS represents an element of volume, r its distance from the point for which we want to know the values of P, Q at time t, and $[E_e]$, $[H_e]$ the impressed forces such as they are in dS at the instant $t - \frac{r}{c}$.

Note 7, § 28

An interesting example is that of two equal electrons moving with constant velocity along a circle in such a way that their positions are continually opposite. We shall suppose that the circle lies in the plane XOY, the center being at the origin of coördinates, and we shall calculate the potentials ϕ and A for a definite time t and for a point $P_{(x, y, z)}$, whose distance r from O is so great that in the expressions for ϕ and A all terms which diminish proportionally to $1/r^2$, $1/r^3$, etc. may be neglected.

This leads to a great simplification when ϕ, A_x, A_y, A_z have to be differentiated. The derivatives with respect to any direction at right angles to OP may be neglected, and the functions may be considered as constant when t and r are changed simultaneously, the one by dt and the other by $dr = c\,dt$. This implies that

$$\frac{\partial}{\partial r} = -\frac{1}{c}\frac{\partial}{\partial t}.$$

Let a be the radius of the circle and ω the angular velocity. Then the potentials may be developed in series according to ascending powers of

$$\beta = \frac{a\omega}{c}, \tag{332}$$

and this quantity, the ratio between the velocity of the electrons and the speed of light, will be taken to be so small that only a few terms of the series need be used. We shall confine ourselves to terms that lead to a factor ω^6 in the final result for the radiation of energy. We shall further simplify by supposing the point P to lie in the zx-plane.

We may begin by considering a single electron. If this particle were placed at O, the effective time would be $t - \frac{r}{c}$. In reality it will be somewhat different from this, say $t - \frac{r}{c} + \tau$. Let the position of the electron be determined by the angle θ, so that the coördinates are

$$x' = a\cos\theta, \quad y' = a\sin\theta,$$

and let θ_0 and θ_e be the values of this angle at the instants $t - \frac{r}{c}$ and $t - \frac{r}{c} + \tau$. Then, in the effective position, the square of the distance r_e from P will be

$$r_e^2 = r^2 - 2\,ax \cos \theta_e + a^2,$$

and as this must be equal to $c^2\left(\frac{r}{c} - \tau\right)^2$, we have the following equation for the determination of τ:

$$-2\,cr\tau + c^2\tau^2 = -2\,ax \cos \theta_e + a^2 = -2\,ax \cos (\theta_0 + \omega\tau) + a^2.$$

With a view to the approximation which we require we may put

$$\tau = \frac{ax}{cr} \cos \theta_0.$$

Thus
$$\theta_e = \theta_0 + \omega\tau = \theta_0 + \beta \frac{x}{r} \cos \theta_0,$$

$$\cos \theta_e = \cos \theta_0 - \beta \frac{x}{r} \sin \theta_0 \cos \theta_0,$$

$$\sin \theta_e = \sin \theta_0 + \beta \frac{x}{r} \cos^2 \theta_0. \tag{333}$$

The direction constants of the line drawn from the electron in its effective position to P are

$$\frac{x - a \cos \theta_e}{r_e}, \quad -\frac{a \sin \theta_e}{r_e}, \quad \frac{z}{r_e},$$

and since the components of the velocity are

$$-a\omega \sin \theta_e, \quad a\omega \cos \theta_e, \quad 0,$$

we find
$$v_r = -\frac{a\omega x}{r_e} \sin \theta_e.$$

In the denominator in the expression for the scalar potential we may replace r_e by r, so that

$$\phi = \frac{e}{4\,\pi r\left(1 + \beta \frac{x}{r} \sin \theta_e\right)},$$

and if we use (333),

$$\phi = \frac{e}{4\,\pi r}\left(1 - \beta \frac{x}{r} \sin \theta_0 - \beta^2 \frac{x^2}{r^2} \cos 2\,\theta_0\right).$$

If, now, we add the second electron, for which at the time $t - \frac{r}{c}$ the position angle is $\theta_0 + \pi$, we find for the total value of the scalar potential

$$\phi = -\gamma \frac{x^2}{r^2} \cos 2\,\theta_0,$$

where
$$\gamma = \frac{e}{2\,\pi r}\,\beta^2. \tag{334}$$

By a similar reasoning the components of the vector potential are found to be

$$A_x = -\gamma \frac{x}{r} \cos 2\,\theta_0, \quad A_y = -\gamma \frac{x}{r} \sin 2\,\theta_0, \quad A_z = 0.$$

We can now simplify by turning the axes OZ and OX in their plane so that the former coincides with OP. Denoting the new coördinates by z' and x', and the angle ZOZ' by α, we have

$$\phi = -\gamma \sin^2 \alpha \cos 2\,\theta_0.$$

$$A_{x'} = -\gamma \sin \alpha \cos \alpha \cos 2\,\theta_0, \quad A_y = -\gamma \sin \alpha \sin 2\,\theta_0,$$

$$A_{z'} = -\gamma \sin^2 \alpha \cos 2\,\theta_0.$$

We have to substitute these values in

$$E_{x'} = -\frac{\partial \phi}{\partial x'} - \frac{1}{c}\dot{A}_{x'}, \quad H_{x'} = \frac{\partial A_{z'}}{\partial y} - \frac{\partial A_y}{\partial z'}, \text{ etc.}$$

But according to what has been said in the beginning of this note the derivatives with respect to x' and y need not be considered, and $\frac{\partial}{\partial z'} = -\frac{1}{c}\frac{\partial}{\partial t}$. Moreover, $\frac{\partial \theta_0}{\partial t} = \omega$, because θ_0 is the angle determining the position of the first electron at the time $t - \frac{r}{c}$. Hence the components of **E** and **H** become

$$E_{x'} = -\frac{2\,\gamma\omega}{c} \sin \alpha \cos \alpha \sin 2\,\theta_0, \quad E_y = \frac{2\,\gamma\omega}{c} \sin \alpha \cos 2\,\theta_0, \quad E_{z'} = 0.$$

$$H_{x'} = -\frac{2\,\gamma\omega}{c} \sin \alpha \cos 2\,\theta_0, \quad H_y = -\frac{2\,\gamma\omega}{c} \sin \alpha \cos \alpha \sin 2\,\theta_0, \quad H_{z'} = 0.$$

Since $2\,\theta_0 = 2\,\omega\left(t - \frac{r}{c}\right) + \text{const.}$, these formulæ represent vibrations propagated in the direction OP and having the frequency $2\,\omega$, as was to be expected. The components $E_{x'}$ and H_y together constitute a rectilinearly polarized system of waves, and so do E_y and $H_{x'}$, and as there is a difference in phase of a quarter period the

resulting waves are elliptically polarized. The flux of energy is in the direction OP and is measured by

$$c(E_{x'}H_y - E_yH_{x'}) = \frac{4\gamma^2\omega^2}{c}(\sin^2\alpha\cos^2\alpha\sin^2 2\,\theta_0 + \sin^2\alpha\cos^2 2\,\theta_0).$$

Over a long lapse of time the mean values of $\sin^2 2\,\theta_0$ and $\cos^2 2\,\theta_0$ are both $\frac{1}{2}$, so that, using also (332) and (334), we may take for the flow of energy per unit of time and unit of area

$$\frac{a^4e^2\omega^6}{2\,\pi^2c^5r^2}(\sin^2\alpha\cos^2\alpha + \sin^2\alpha).$$

Multiplying this by $2\,\pi r^2 \sin\alpha\, d\alpha$ and integrating with respect to α between the limits 0 and π, we find for the total radiation of energy per unit of time

$$\frac{2\,a^4e^2\omega^6}{5\,\pi c^5}.$$

This can be verified by attending to the forces acting on the electrons in their circular motion. On account of its own field each electron experiences a force opposite to the direction of its motion whose magnitude, according to formula (83), is

$$\frac{ae^2\omega^3}{6\,\pi c^3},$$

and in addition to this there is a force due to the field produced by the other particle. When this force has been determined, we know the forces that must be applied to the particles in order to maintain their motion, and the work of these forces must be equivalent to the radiated energy.

The above result for the radiation may be compared with the corresponding formula for a single electron moving in a circle. The field which this produces at a great distance is the result of the superposition of two fields that correspond to rectilinear vibrations, at right angles to each other, into which the circular motion can be decomposed, and in this superposition the mean flows of energy are simply added. Hence, by virtue of (81), the energy radiated per unit of time is given by

$$\frac{a^2e^2\omega^4}{6\,\pi c^3},$$

a value that can also be deduced from the formulæ of this note. It may be added that if instead of two electrons we had three, four, or more of them, placed at equal distances on the circle and moving

with a constant angular velocity, the term with ω^6 would disappear.* It would be necessary to go to higher powers of β in the series which we have used, and so for n electrons the radiated energy will be represented by an expression containing the factor

$$\frac{\omega^2}{c}\left(\frac{a\omega}{c}\right)^{2n}.$$

Note 8, § 28

Let σ be any closed surface surrounding a system of moving electric charges which produce an electromagnetic field. By a properly chosen distribution of electromotive forces working in tangential directions in this surface, or rather in a layer of infinitely small thickness along it, the external field may be made to disappear without any change of the internal field.

We shall distinguish by the suffixes 1 and 2 the values of quantities just inside and outside the layer, and we shall draw the normal n toward the outside. Moreover, we shall introduce at any point of σ two tangential directions h and k at right angles to each other, choosing them in such a way that the directions h, k, n may be made, by a common rotation, to coincide with those of x, y, z.

Our fundamental equations have the ordinary form

$$\int H_s ds = \frac{1}{c}\int \dot{D}_n d\sigma, \quad \int E_s ds = -\frac{1}{c}\int \dot{B}_n d\sigma, \tag{335}$$

but we have now to distinguish, at least within the thickness of the layer, between $\mathbf{D}$ and $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{H}$, having

$$\mathbf{D} = \mathbf{E} + \mathbf{E}_e, \quad \mathbf{B} = \mathbf{H} + \mathbf{H}_e. \tag{336}$$

Applying (335) to a contour composed of two sides having the direction h (or that of k) and situated just outside and just inside the layer, and two other sides, infinitely small in comparison with the first two and crossing the layer in the direction of the normal, we find

$$H_{h2} - H_{h1} = \frac{1}{c} C_k, \quad H_{k2} - H_{k1} = -\frac{1}{c} C_h, \tag{337}$$

$$E_{h2} - E_{h1} = -\frac{1}{c} M_k, \quad E_{k2} - E_{k1} = \frac{1}{c} M_h, \tag{338}$$

where
$$C_h = \int \dot{D}_h dn, \quad C_k = \int \dot{D}_k dn, \tag{339}$$

$$M_h = \int \dot{B}_h dn, \quad M_k = \int \dot{B}_k dn. \tag{340}$$

* For additional information on this question see Larmor, *Phil. Mag.*, Vol. 42 (1921), p. 595; Leigh Page, *Phys. Review*, Vol. 20 (1922), p. 18. — Ed.

In these expressions the positive quantity dn is a line-element along the normal, and the integrals extend over the thickness δ of the layer. They represent the components of what may be called the electric current **C** and the magnetic current **M** in the layer or in the surface.

When δ is made to approach the limit zero, these currents may still have finite magnitudes; for this it is only necessary that the values of D_h, D_k, B_h, and B_k increase sufficiently. In the limiting case, equations (337) and (338) are the only conditions that have to be fulfilled at the surface. They show that finite values of **E** and **H** on both sides of the surface are quite consistent with finite currents in it, and we may therefore imagine that **E** and **H** are finite also within the thickness of the layer.

In this case, if in (339) and (340) we substitute the values (336), the terms with $\dot{\mathbf{E}}$ and $\dot{\mathbf{H}}$ will contribute nothing to the integrals, and these can have finite values only if the electromotive and magnetomotive forces are made to increase indefinitely. Thus, if we put

$$\int E_{eh}dn = P_h, \quad \int E_{ek}dn = P_k,$$

$$\int H_{eh}dn = Q_h, \quad \int H_{ek}dn = Q_k,$$

the conditions (337) and (338) become

$$H_{h2} - H_{h1} = \frac{1}{c}\dot{P}_k, \qquad H_{k2} - H_{k1} = -\frac{1}{c}\dot{P}_h, \tag{341}$$

$$E_{h2} - E_{h1} = -\frac{1}{c}\dot{Q}_k, \quad E_{k2} - E_{k1} = \frac{1}{c}\dot{Q}_h. \tag{342}$$

The quantities P_h, P_k, Q_h, Q_k may be called the components of the electromotive and the magnetomotive action in the surface. We shall now choose them so that we may obtain the result that was announced at the beginning of this note.

In the original state of things, when no electromotive and magnetomotive forces were applied, the field belonging to the moving charges extended to infinite distance without any breach of continuity at the surface σ. Let E_h, E_k, H_h, H_k be the components of E and H existing at the surface in this field. Then, if we choose the impressed actions in such a manner that

$$\left.\begin{array}{ll} \dot{P}_h = cH_k, & \dot{P}_k = -cH_h \\ \dot{Q}_h = -cE_k, & \dot{Q}_k = cE_h \end{array}\right\}, \tag{343}$$

all conditions are satisfied when in the external space there is no field at all, whereas in the internal space all has remained as it was.

Indeed, the conditions (341) and (342) are satisfied by the values (343) if the components of **H** and **E** with the suffix 2 are zero and those with the index 1 have the original values.

It should be noticed that the choice defined by (343) can always be made, and that when the vectors **E** and **H** are known, we always obtain by these formulæ the same vectors $\dot{\mathbf{P}}$ and $\dot{\mathbf{Q}}$ in the surface, whatever be the choice of the directions h and k.

We can express in words the meaning of (343) by saying that the vectors $\frac{1}{c}\dot{\mathbf{P}}$ and $\frac{1}{c}\dot{\mathbf{Q}}$ have the same magnitudes as the total tangential components, say $\overline{\mathbf{H}}$ and $\overline{\mathbf{E}}$, of **H** and **E**, and are at right angles to these components, the rotation from $\overline{\mathbf{H}}$ to $\frac{1}{c}\dot{\mathbf{P}}$ over 90° being clockwise and that from $\overline{\mathbf{E}}$ to $\frac{1}{c}\dot{\mathbf{Q}}$ counterclockwise, when viewed from the outside.

Moreover, it is interesting to show that in the new state of things, notwithstanding the vanishing of the outside field, both the electric and the magnetic current are solenoidally distributed.

Let s be a closed line in the surface, and ν a line tangential to the surface and normal to s, directed toward the outside. We shall choose the positive direction along s in such a way that ν and s may be taken for h and k; then the motion along the line in the positive direction will correspond to the direction of the normal n.

If now we consider two surfaces σ_1 and σ_2 infinitely near the layer σ, the one on the inside of it and the other on the outside, and a third surface passing through s and intersecting the surface σ at right angles, these three will limit a certain space, and the total surface integral of the electric current across them must be zero. At the points of σ_2 there is no current at all, so that the condition becomes

$$-\int \dot{E}_{n1} d\sigma + \int C_\nu ds = 0,$$

and this is true because, by virtue of one of Maxwell's equations, the first integral has the value

$$c\int H_{s1} ds$$

and the second has the same value, because according to one of the equations (337) we have

$$-\frac{1}{c} C_\nu = H_{s2} - H_{s1}$$

and $H_{s2} = 0$.

The solenoidal distribution of the magnetic current can be verified in a similar way.

It ought to be remarked that the method fails in the case of a constant electric or magnetic field, for then, as is seen from

(343), we should have to imagine constantly increasing impressed forces. But if the given field is variable and, say, periodic, we can decompose the values of H and E at the surface by means of Fourier's theorem, and all the periodic partial fields which we find in this way can be got rid of by suitable electromotive and magnetomotive forces.

Note 9, § 32

In expositions of the theory of relativity we often imagine (and I have done so in the text) different observers, each of whom has a definite position in one of the systems of reference. It must, however, be remarked that this mode of presenting the subject, though it may be in the interest of clearness, is not at all essential and may therefore lead to misunderstandings.

The only really important thing is the choice of the system of reference that we wish to use in the discussion of a problem. In choosing it we need not in the least think of our own state of rest or motion. Without leaving his chair an astronomer makes his calculations in systems of coördinates in which the earth has any velocity he desires.

Thus, for the discussion contained in § 32 it would have been wholly sufficient to imagine two systems of measuring rods and clocks, the ones R, C being at rest in the system x, y, z, t, and the others, R', C', moving in it, so that they are at rest in the system x', y', z', t'. The comparisons between the rods R and R', and between the clocks C and C', can be made in different ways. If we fix our attention on a rod R' and on the points A and B of a rod R with which its extremities coincide at moments so chosen that two clocks C placed at A and B indicate the same t, we shall find R' to be shorter than the rod R. Conversely, we shall find R to be shorter than R' when we read on the rod R' the positions A' and B' of the ends of R at instants when the hands of two clocks C' placed at A' and B' have the same position.

Similarly, the clocks can be compared in various ways. We may, in the first place, using the system x, t, consider the motion of the clocks C', moving with the velocity v. Let C' be one of these clocks, and let C_1, C_2, C_3, $\cdots$ be clocks of the other system near which it passes successively. If, at the instants of meeting, the indications of these clocks are t_1, t_2, t_3, $\cdots$, and if t'_1, t'_2, t'_3, $\cdots$ are the indications of C' itself at these instants, the differences $t'_2 - t'_1$, $t'_3 - t'_2$, $\cdots$ will be smaller than $t_2 - t_1$, $t_3 - t_2$, $\cdots$ in the ratio given by $\sqrt{1 - \frac{v^2}{c^2}}$.

We therefore say that in the system x, t the clocks C' go more slowly than the stationary ones C.

Exactly in the same way we are led to the conclusion that in the system of reference x', t' the clocks C go more slowly than C'. We can ascertain it by keeping in view a definite clock C, comparing it with the different clocks of the other group, C'_1, C'_2, C'_3, $\cdots$. Let t'_1, t'_2, t'_3, $\cdots$ be the indications of these clocks at the instants when C is near each of them (so that these symbols $t'_1, t'_2, t'_3 \cdots$ now have meanings different from those which they recently had), and let t_1, t_2, t_3, $\cdots$ be the indications of C at these same instants; then the differences $t_2 - t_1$, $t_3 - t_2$, $\cdots$ will be smaller than the differences $t'_2 - t'_1$, $t'_3 - t'_2$, $\cdots$.

It should be noted that the observations which we have been considering consist simply in the establishment of coincidences between lines of the scales R and R' and the comparison of two clocks very close to each other. It may be taken for granted that all this may be done with exactly the same results by any observer, whatever be his position and his state of rest or motion in the system x, t, or x', t'.

It remains to say some words about the question raised at the end of § 32. If, for the sake of brevity, we leave out of account the changes produced in the periods that were called τ_1, τ_3, and τ_5, on the ground, if necessary, that the times τ_2 and τ_4 are so long that the changes taking place in them far predominate, the theory of relativity surely requires that if one clock A has a fixed position in the system x, t, and if a second clock B, first near A, is made to move along the rod R, say to the right, over a certain distance, and then back again with the same velocity, it will be found to have lagged behind relatively to A. If at the moment of departure the indications of the two agree, $t_B = t_A$, we shall have at the end of the journey $t_B < t_A$. The same effect will occur when the motion of B is first to the left and then to the right.

We shall denote either of these cases, in which A is at rest while B moves, as case I, and we shall compare with it another experiment in which, always in the system x, t, the clock B is at rest, whereas A moves to and fro over a certain distance on one side or the other of the position of B. In this case II we shall have at the end $t_A < t_B$, and if otherwise the circumstances are the same in the two cases, $t_B - t_A$ will be now what $t_A - t_B$ was formerly.

In connection with this it may be remarked that one must be careful in formulating " principles of relativity " and in considering them as more or less self-evident. Indeed, one might feel tempted to admit as a postulate that in the case of the two clocks moving in the

system x, t along a line parallel to OX, all will depend on the way in which the mutual distance changes in course of time. If this were so, the two cases I and II ought to be considered as identical. There would be nothing to distinguish between one clock and the other, and so there could be no reason why, in the end, the indications t_A and t_B should be different. Thus it appears that this postulate of "relative distance," as we may call it for a moment, is in contradiction with Einstein's theory of relativity; it cannot be true when we consider this theory as verified by the observed phenomena.

Of course, if we imagine a stationary ether, we can understand why the principle of relative distance fails; the final state of things may then very well depend on whether it has been the clock A or the clock B that has moved through the medium.

Note 10, § 40

The principles of conservation of energy and of momentum have recently been applied to a collision between an electron and a light-quantum, and a very beautiful theory of the change in frequency of the light-quantum has been developed in this way by A. H. Compton (*Phys. Review*, Vol. 21 (1923), pp. 207, 483; Vol. 23 (1924), p. 439) and P. Debye (*Phys. Zeitschr.*, Bd. 24 (1923), p. 161).

The principles may be applied to any two particles which approach and fly apart if the mutual energy is initially and finally zero and if there is no loss of energy and momentum by radiation during the encounter.

Let the initial momenta of the particles be represented in magnitude and direction by $c \cdot AP$ and $c \cdot PB$ respectively, AP and PB being two lines; then the total momentum is represented in magnitude and direction by $c \cdot AB$.

Now let the final momenta be represented in magnitude and direction by $c \cdot AQ$ and $c \cdot QB$ respectively, the total momentum being the same as before. According to the two formulæ for momentum G and energy E we have the relation

$$\sqrt{\frac{G^2}{c^2} + m^2} = \frac{E}{c^2},$$

where m, the "mass," is zero for a light-quantum,* so that the conservation of energy may be expressed by saying that the sum of the quantities

$$\sqrt{\frac{G^2}{c^2} + m^2}$$

remains unaltered in the encounter.

* This results from the formula for the energy (§ 39).

Thus if, before the encounter,

$$\sqrt{AP^2 + m_1^2} + \sqrt{PB^2 + m_2^2}$$

has a certain value μ, we shall have, after the encounter,

$$\sqrt{AQ^2 + m_1^2} + \sqrt{QB^2 + m_2^2} = \mu.$$

To interpret this equation we begin with the case in which Q lies in the plane APB. The locus of Q is then an ellipse determined by the intersection of the plane APB and a prolate spheroid whose foci are situated on the perpendiculars to the plane erected at A and B at distances m_1 and m_2 from these points. This ellipse will have AB as axis (but not A and B as foci unless $m_1 = m_2 = 0$). If $AP = 0$ it has A as vertex. Revolving this ellipse about the line AB, we obtain a spheroid which is the locus of the point Q in space. When the total momentum is zero A and B coincide and the axis of the first spheroid is perpendicular to the plane APB. The ellipse then becomes a circle and the second spheroid a sphere. We thus have W. Pauli's result that in this case the momentum of each particle is unaltered in magnitude by a collision. — Ed.

Note 11, § 44

Let x, y, z be the coördinates and $\dot{x}, \dot{y}, \dot{z}$ the components of the velocity $\mathbf{v}$ of a particle, m the mass, and X, Y, Z the components of the force acting on it. Then, since the components of the momentum are

$$\frac{m\dot{x}}{\sqrt{1-\frac{v^2}{c^2}}} \text{ etc.,}$$

the equations of motion are

$$X = \frac{d}{dt}\frac{m\dot{x}}{\sqrt{1-\frac{v^2}{c^2}}} \text{ etc.};$$

and we have

$$Xx = x\frac{d}{dt}\frac{m\dot{x}}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{d}{dt}\frac{mx\dot{x}}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{m\dot{x}^2}{\sqrt{1-\frac{v^2}{c^2}}},$$

with similar formulæ for Yy and Zz. Hence, if the virial is still defined by

$$V = \sum(Xx + Yy + Zz),$$

$$V = \frac{d}{dt}\sum\frac{m(x\dot{x} + y\dot{y} + z\dot{z})}{\sqrt{1-\frac{v^2}{c^2}}} - \sum\frac{mv^2}{\sqrt{1-\frac{v^2}{c^2}}}.$$

The first term is of the form $d\psi/dt$, and for its mean value for an interval ranging from t_1 to t_2 we may write

$$\frac{\psi_2 - \psi_1}{t_2 - t_1}.$$

If the state of motion is stationary, so that there are no changes going on in the long run in a predominating direction, the difference $\psi_2 - \psi_1$ will not increase proportionally to $t_2 - t_1$ when the length of the interval is augmented, and the mean value of $d\psi/dt$ for a long time will be zero.

Note 12, § 44

According to the theorem proved in the preceding note, $-2\,\overline{T'}$ will be equal to the mean virial of all forces acting on the masses; that is,

$$\overline{V} + \overline{V}_e = -2\,\overline{T'}$$

if V_e is the virial of the forces due to the electromagnetic field. In the ordinary notations the force acting on the charge contained in an element of volume dS is

$$\rho\left(\mathbf{E} + \frac{1}{c}\,[\mathbf{v}\cdot\mathbf{H}]\right)dS,$$

and if $\mathbf{r}$ is the radius vector from the origin to the point considered, we have

$$V_e = \int \rho\left(\mathbf{r}\cdot\left\{\mathbf{E} + \frac{1}{c}\,[\mathbf{v}\cdot\mathbf{H}]\right\}\right)dS. \tag{344}$$

This is seen by remarking that the expression $Xx + Yy + Zz$ is the scalar product of $\mathbf{r}$ and the force (X, Y, Z).

We shall now transform the above expression, supposing for the sake of simplicity that ρ, $\mathbf{v}$, $\mathbf{E}$, and $\mathbf{H}$ are continuous functions of the coördinates. This saves us the trouble of considering integrals over surfaces of discontinuity and does not diminish the generality of the proof, because abrupt changes can be treated as limiting cases of gradual ones.

The integration in (344) extends to all points where there is an electromagnetic field. We can, however, first confine it to the space within some closed surface. The partial integrations which we shall have to perform will then give rise to integrals extended over this surface. We shall, however, omit these, supposing either that the system is of limited extent and that at the surrounding surface σ there is no field, or that, though the system extends to infinity, the forces $\mathbf{E}$ and $\mathbf{H}$ decrease so rapidly that the integrals in question tend to zero when the surface is made to recede to infinity on all sides.

By the substitution of div $\mathbf{E}$ for ρ and of c curl $\mathbf{H} - \dot{\mathbf{E}}$ for $\rho\mathbf{v}$, the virial V_e is decomposed into three parts which we shall consider successively.

$$V_{e1} = \int (\mathbf{r} \cdot \mathbf{E}) \operatorname{div} \mathbf{E} \, dS = \int \left[(xE_x + yE_y + zE_z) \frac{\partial E_x}{\partial x} + \cdots \right] dS,$$

where $\cdots$ means that the expression that has been written out has to be followed by two similar ones that are derived from it by cyclic permutation of x, y, z.

Integrating by parts and using relations like

$$\frac{\partial E_y}{\partial x} = \frac{\partial E_x}{\partial y} - \frac{1}{c} \dot{H}_z,$$

drawn from one of Maxwell's equations,

$$\begin{aligned}
V_{e1} &= -\int \left[\frac{1}{2} E_x^2 + \cdots + E_x \left(y \frac{\partial E_y}{\partial x} + z \frac{\partial E_z}{\partial x} \right) + \cdots \right] dS \\
&= -\int \frac{1}{2} E^2 dS - \int \left[E_x \left(y \frac{\partial E_x}{\partial y} + z \frac{\partial E_x}{\partial z} \right) + \cdots \right] dS \\
&\qquad + \frac{1}{c} \int [E_x(y\dot{H}_z - z\dot{H}_y) + \cdots] dS \\
&= -\int \frac{1}{2} E^2 dS + \int \left[\left(\frac{1}{2} E_x^2 + \frac{1}{2} E_x^2 \right) + \cdots \right] dS \\
&\qquad - \frac{1}{c} \int [x(E_y\dot{H}_z - E_z\dot{H}_y) + \cdots] dS \\
&= \int \frac{1}{2} E^2 dS - \frac{1}{c} \int (\mathbf{r} \cdot [\mathbf{E} \cdot \dot{\mathbf{H}}]) \, dS. \qquad (345)
\end{aligned}$$

The second part is

$$\begin{aligned}
V_{e2} &= \int (\mathbf{r} \cdot [\operatorname{curl} \mathbf{H} \cdot \mathbf{H}]) \, dS \\
&= \int \left[x \left\{ \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \right) H_z - \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right) H_y \right\} + \cdots \right] dS \\
&= -\int \left[xH_x \left(\frac{\partial H_z}{\partial z} + \frac{\partial H_y}{\partial y} \right) + \cdots \right] dS + \int \left[\frac{1}{2} (H_z^2 + H_y^2) + \cdots \right] dS.
\end{aligned}$$

Since div $\mathbf{H} = 0$, the first integral becomes

$$-\int [\tfrac{1}{2} H_x^2 + \cdots] dS,$$

so that

$$V_{e2} = \int \tfrac{1}{2} H^2 dS.$$

Finally

$$V_{e3} = -\frac{1}{c} \int (\mathbf{r} \cdot [\dot{\mathbf{E}} \cdot \mathbf{H}]) \, dS,$$

which, combined with the last term of (345) gives

$$-\frac{1}{c} \frac{d}{dt} \int (\mathbf{r} \cdot [\mathbf{E} \cdot \mathbf{H}]) \, dS,$$

the mean value of which over a long time is zero when the system is in a stationary state. Thus, if E is the electromagnetic energy,

$$\overline{V}_e = \overline{E},$$

from which the formula given in the lecture follows directly.

Note 13, § 45

In order to prove what has been said about the sign of the moduli of elasticity we shall first recall some propositions of the theory of elasticity.

Let x, y, z be the coördinates of a point of an elastic body in its natural state of equilibrium, and ξ, η, ζ the components of a small displacement varying from point to point. The deformation of the body will be specified by the following six components, three dilatations and three shearing strains:

$$x_x = \frac{\partial \xi}{\partial x}, \quad y_y = \frac{\partial \eta}{\partial y}, \quad z_z = \frac{\partial \zeta}{\partial z},$$

$$x_y = y_x = \frac{\partial \xi}{\partial y} + \frac{\partial \eta}{\partial x}, \quad y_z = z_y = \frac{\partial \eta}{\partial z} + \frac{\partial \zeta}{\partial y}, \quad z_x = x_z = \frac{\partial \zeta}{\partial x} + \frac{\partial \xi}{\partial z};$$

and the potential energy U per unit of volume will be a function of these quantities. Differentiating it, we obtain the values of the normal and the tangential stresses,

$$X_x = \frac{\partial U}{\partial x_x} \text{ etc.,} \quad X_y = \frac{\partial U}{\partial x_y} \text{ etc.}$$

The potential energy may be taken to be zero in the natural state, and if in that state the body is free from external forces, so that we may suppose that there are no internal stresses, the first derivatives of U must vanish for $x_x = 0$ etc., $x_y = 0$ etc. Hence for small displacements U will be a homogeneous function of the second degree of the components of strain. For a body having the symmetry of a cubical crystal this function can be shown to have the form

$$U = \tfrac{1}{2} A(x_x^2 + y_y^2 + z_z^2) + \tfrac{1}{2} B(x_x + y_y + z_z)^2 + \tfrac{1}{2} C(x_y^2 + y_z^2 + z_x^2), \quad (346)$$

when the axes of coördinates coincide with the principal axes of the crystal. The constants A, B, C may be called moduli of elasticity.

It will be possible to calculate their values when we know the forces acting between the particles. We shall suppose these to be small rigid bodies, in each of which we distinguish a center; for

example, at the place of the nucleus of an atom. By turning about their centers the particles may take different orientations, but we shall begin by supposing that all displacements ξ, η, ζ consist in pure translations, the orientation of each particle being kept unaltered by a proper constraint. On this assumption the relative position of two particles P and Q is wholly determined by the relative coördinates x, y, z of Q with respect to P, by which we mean the relative coördinates of the center of Q with respect to that of P. The mutual potential energy ϕ of the two particles P and Q will be a definite function of x, y, z; but this function will not necessarily have the same form for different pairs, for in a pair P', Q' the relative orientation of the particles need not be the same as it is in P, Q. Even the constitution of two pairs may be different; P, Q may consist of two ions of sodium, and P', Q' of two ions of chlorine or of one of sodium and one of chlorine. At all events we shall find the total potential energy by taking the sum of the values of ϕ for all pairs of particles. We may also say that when $\Delta\phi$ is the change in a particular ϕ, produced by the displacements ξ, η, ζ, the potential energy is given by

$$\sum \Delta\phi.$$

The calculation of sums of this kind is greatly simplified by the fact that the interaction between two elements of volume is made up of a great number of electric attractions and repulsions, which practically counterbalance each other as soon as the distance is a somewhat high multiple of that between neighboring molecules. So, in this theory, just as in the old theory of elasticity, though for a different reason, there is a sensible interaction only at distances that are very small in comparison with the dimensions of the bodies on which we make our observations. Hence by far the greater part of the terms in $\Sigma\Delta\phi$ are due to pairs of particles lying in the interior of the body, and in the case of a homogeneous strain the sum may be taken to be proportional to the volume of the body. In what follows we shall refer it to unit of space.

With these simplifications the moduli are easily calculated. Take, in the first place, a dilatation in the direction of x, in which x_x has the value q. The relative coördinates of one particle of a pair with respect to the other, which first were x, y, z, are changed to $(1+q)x$, y, z, and when we confine ourselves to terms of the second order we may write

$$\Delta\phi = qx\frac{\partial\phi}{\partial x} + \frac{1}{2}q^2x^2\frac{\partial^2\phi}{\partial x^2}.$$

On the ground of what has been said already the sum of the first term must vanish. Therefore

$$U = \frac{1}{2} q^2 \sum x^2 \frac{\partial^2 \phi}{\partial x^2};$$

and if this is compared with (346),

$$A + B = \sum x^2 \frac{\partial^2 \phi}{\partial x^2}.$$

Of course, on account of the symmetry in the structure, the sums

$$\sum y^2 \frac{\partial^2 \phi}{\partial y^2} \quad \text{and} \quad \sum z^2 \frac{\partial^2 \phi}{\partial z^2}$$

will have the same value.

Next consider a shearing strain in which the particles are displaced in the direction y over distances proportional to x. By this the relative coördinates x, y, z are changed to x, $y + qx$, z, and we have $x_y = y_x = q$,

$$\Delta\phi = qx \frac{\partial \phi}{\partial y} + \frac{1}{2} q^2 x^2 \frac{\partial^2 \phi}{\partial y^2},$$

which leads us to

$$C = \sum x^2 \frac{\partial^2 \phi}{\partial y^2}.$$

The sum $\sum x^2 \frac{\partial^2 \phi}{\partial z^2}$ has the same value, and we have therefore

$$A + B + 2\,C = \sum x^2 \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \right).$$

Now this sum must necessarily be zero. This is seen by reverting to a pair of particles P, Q. If we place the origin of coördinates at the center of P, so that the coördinates of Q's center become x, y, z, and if we fix our attention on a charge e in P and a charge e' in Q, lying at points whose coördinates with respect to the center of P and that of Q, respectively, are a, b, c and a', b', c', we have

$$\phi = \sum \frac{ee'}{r},$$

where

$$r = \sqrt{(x + a' - a)^2 + (y + b' - b)^2 + (z + c' - c)^2},$$

and where the sum is extended to all combinations of a charge in P with a charge in Q. According to our assumptions, a, b, c, a', b', c' are to be considered as constants when ϕ is differentiated, so that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \frac{1}{r} = 0$$

and

$$A + B + 2\,C = 0,$$

showing that the moduli cannot all be positive. There must be values of x_x etc. for which U becomes negative, and this proves that the equilibrium in the original state is unstable.

This conclusion is not invalidated by our assumption that a suitable constraint prevents the molecules from turning. In fact, let us start from the original state of the body with the orientations Ω naturally existing in it, and let us then, maintaining these orientations, apply a homogeneous strain of such a kind as to make the potential energy less than it was. With the new positions of the centers of the molecules the orientations Ω are in all probability not those which they would take under the action of the mutual forces. Thus, if we do away with the constraints, the molecules will pass to other orientations Ω'. This will certainly be accompanied by a new diminution of the potential energy, and we may therefore be sure that it would also have decreased if, from the outset, the particles had been free to turn.

It was not necessary for our purpose to calculate A, B, and C separately. If, however, we want to do so, we have only to apply to a third case the reasoning that served us for two specially chosen deformations. Thus, by considering a dilatation equal in all directions one finds

$$A + 3\,B = \sum x^2 \frac{\partial^2 \phi}{\partial x^2} + 2 \sum xy \frac{\partial^2 \phi}{\partial x \, \partial y}.$$

Note 14, § 52

A few words may be said here about the fluctuations in the amount of energy contained in a space filled with the black radiation that is emitted by surrounding bodies kept at a given temperature. A similar problem concerning the distribution of molecules may serve as an introduction. Take the case of a gas consisting of n molecules, which we may suppose to be so small that they perform their heat motion without any mutual collisions. If, by an imaginary plane, we divide the total volume v into two, say, unequal parts v_1 and v_2, what will be the partition of the gas between these parts?

We shall suppose (and this is a natural assumption) that of the path L traveled over by a molecule during a lapse of time in which there are very numerous reflections by the walls the part

$$\frac{v_1}{v} L = w_1 L$$

lies in v_1 and the part

$$\frac{v_2}{v} L = w_2 L$$

in v_2. Starting from this, we are led to the following conclusions:

The probability that at a definite instant an individual molecule will be in v_1 is given by w_1, and the probability that it will lie in v_2 is w_2.

The probability that n_1 specially designed molecules will all lie in v_1 and the n_2 remaining ones in v_2 is

$$w_1^{n_1} w_2^{n_2}.$$

As, with the n molecules, two groups, the one with n_1 particles and the other with n_2 particles, may be formed in

$$\frac{n!}{n_1!\,n_2!}$$

different ways, the probability of a partition n_1, n_2, whichever be the n_1 molecules that lie in v_1, is

$$\frac{n!}{n_1!\,n_2!} w_1^{n_1} w_2^{n_2} = \frac{n!}{n_1!\,(n-n_1)!} w_1^{n_1} w_2^{n-n_1}. \tag{347}$$

This may be applied to the cases $n_1 = 0$, $n_1 = n$ if the coefficient is given the value 1.

The above expression is one term in the expansion of

$$(w_1 + w_2)^n,$$

and we have, therefore, since $w_1 + w_2 = 1$,

$$\sum_{n_1=0}^{n_1=n} \frac{n!}{n_1!\,(n-n_1)!} w_1^{n_1} w_2^{n-n_1} = 1. \tag{348}$$

This means that the sum of all the probabilities in question is 1, as obviously it must be.

Of course, (348) holds for any value of n. We may, for example, replace n by $n-1$ or $n-2$.

We shall suppose n to be a very large number. Now when we go over the numerous terms in the expansion of

$$(w_1 + w_2)^n,$$

we shall find them first to increase and then to diminish. The place of the maximum is determined by the condition that the ratio between two successive terms is practically equal to unity.

When in (347) n_1 is replaced by $n_1 + 1$, the expression is multiplied by the factor

$$\frac{n - n_1}{n_1 + 1} \frac{w_1}{w_2},$$

for which we may write, if n_1 also is very great,

$$\frac{n - n_1}{n_1} \frac{w_1}{w_2},$$

so that the most probable distribution is the uniform one in which

$$n_1 = w_1 n, \quad n_2 = w_2 n. \tag{349}$$

Distributions different from this are noways excluded, the probability for the *exact* realization of (349) (or of the integer numbers n_1 and n_2 that are nearest to $w_1 n$ and $w_2 n$) being even very small; and if we repeat the experiment many times, or consider the gas at a great number of instants, we may expect to find all imaginable partitions realized, each with a frequency proportional to its probability.

If ϕ is some function of n_1, its mean value in all these cases will be given by

$$\overline{\phi} = \sum_{n_1=0}^{n_1=n} \frac{n!}{n_1!\,(n-n_1)!}\,\phi\, w_1^{n_1} w_2^{n-n_1}.$$

Thus

$$\begin{aligned}\overline{n}_1 &= \sum_{n_1=1}^{n_1=n} \frac{n!}{(n_1-1)!\,(n-n_1)!}\, w_1^{n_1} w_2^{n-n_1} \\ &= n w_1 \sum_{n_1'=0}^{n_1'=n-1} \frac{(n-1)!}{n_1'!\,(n-1-n_1')!}\, w_1^{n_1'} w_2^{n-1-n_1'} = n w_1,\end{aligned}$$

$$\begin{aligned}\overline{n_1(n_1-1)} &= \sum_{n_1=2}^{n_1=n} \frac{n!}{(n_1-2)!\,(n-n_1)!}\, w_1^{n_1} w_2^{n-n_1} \\ &= n(n-1)w_1^2 \sum_{n_1''=0}^{n_1''=n-2} \frac{(n-2)!}{n_1''!\,(n-2-n_1'')!}\, w_1^{n_1''} w_2^{n-2-n_1''} \\ &= n(n-1)w_1^2,\end{aligned}$$

$$\overline{n_1^2} = \overline{n_1(n_1-1)} + \overline{n_1} = n(n-1)w_1^2 + n w_1.$$

Now let

$$\Delta n_1 = n_1 - n w_1$$

be the difference between the most probable or mean value nw_1 and the actual one n_1. By what precedes

$$\overline{\Delta n_1} = \overline{n_1} - nw_1 = 0,$$
$$\overline{(\Delta n_1)^2} = \overline{n_1^2} - 2\,nw_1\overline{n_1} + n^2w_1^2 = n(w_1 - w_1^2) = \bar{n}_1 w_2.$$

In the special case $v_1 = \frac{1}{2}\,v$ we have

$$\overline{(\Delta n_1)^2} = \tfrac{1}{2}\,\overline{n_1},$$

and when v_1 is only a very small part of v,

$$\overline{(\Delta n_1)^2} = \overline{n_1}. \tag{350}$$

By a mode of reasoning based on the consideration of the entropy of the system, and using Planck's black-body radiation formula, Einstein has established a formula for the fluctuations in the energy of radiation. Let v be the volume of a radiation field surrounded by ponderable bodies which are kept at a definite temperature, E the mean or most probable value of the energy contained in the field, so far as that energy belongs to frequencies between ν and $\nu + d\nu$, and ϵ the deviation of the actual value of this energy from the mean value. Then, according to Einstein's theory,

$$\overline{\epsilon^2} = h\nu E + \frac{c^3}{8\,\pi\nu^2 d\nu}\,\frac{E^2}{v}.$$

Of the two terms on the right-hand side the second is easily explained by the classical theory of radiation. Indeed, the radiation field results from the superposition of systems of waves intersecting in all directions. The phases of these waves may be conceived as varying at random, and so the interference may lead to an energy that is somewhat greater or less in one case than in the other. When this is worked out, we find exactly the last term of the equation.

On the other hand, the theory of light-quanta leads to an interpretation of the term $h\nu E$. As the black radiation contained in the field v is only a small part of the total system which comprises the surrounding bodies, we may expect between the mean number of quanta n and the deviation Δn from the mean a relation like (350),

$$\overline{(\Delta n)^2} = n.$$

But $E = h\nu n$, $\epsilon = h\nu\Delta n$. Thus $\overline{\epsilon^2} = h\nu E$.

Is it possible to explain the two terms by one and the same theory? We must certainly give up all hope of doing so with the old views, for these can never make us understand the term $h\nu E$. The theory of

light-quanta, however, is more promising. We might begin by imagining the quanta exactly to reproduce the distribution of energy in a field where different waves interfere (compare Note 16); by this we should have explained the last term in the equation. After that we might suppose that there are slight deviations from this exact reproduction comparable to the deviations from a uniform distribution of a gas, and we might try to account for the first term by means of these. But it seems difficult to form a clear and consistent picture of all this.

That the mean square of the deviation is made up of two parts as the formula shows, may be illustrated to a certain extent by a general theorem. Let there be two causes which make a variable quantity differ from its mean value. The two kinds of deviations, α and β, thus produced, may each of them have many different values, both positive and negative, and we shall suppose α and β to be mutually independent, so that when we speak, for example, of the mean value of α or of α^2, we need not think of the values of β with which a particular α is combined. Then, if $\overline{\alpha}=0$ and $\overline{\beta}=0$, we shall also have $\overline{\alpha\beta}=0$, and the mean square of the actual deviation $\alpha+\beta$ will be

$$\overline{(\alpha+\beta)^2}=\overline{\alpha^2}+\overline{\beta^2}.$$

Note 15, § 52

A reflection of a quantum may be regarded as an absorption and emission taking place simultaneously. If an atom plays a part in each of these processes, the question of the way in which the laws of reflection depend on the properties of the atom needs to be answered. In the theory of normal X-ray reflection which has been developed by W. Duane with the aid of his idea of quantizing linear momentum the properties of the atoms influence the reflection only in an indirect manner, inasmuch as they determine the structure of the crystal. A type of reflection in which a direct influence of the atomic structure appears would be called abnormal.

This theory of normal reflection is based on the idea that quanta move in such a way that where there is light there are quanta and that no quanta go to places where there is darkness. It is interesting to notice that this idea had been formulated by Dr. Lorentz in § 50.

The theory of X-ray reflection has been further developed along these lines by A. H. Compton (*Proc. Nat. Acad. Sci.*, Vol. 9 (1923), p. 359), and the theory of interference has been discussed with the aid of the new ideas of Duane and Compton by P. S. Epstein

and P. Ehrenfest (ibid., Vol. 10 (1924), p. 133). They use the Bohr principle of correspondence as a guide. An interesting exposition of the theory of Duane and Compton has been given by W. P. Davey (*General Electric Review*, Vol. 27 (1924), p. 742). — ED.

Note 16, § 53

We shall briefly discuss here some further questions to which the theory of light-quanta gives rise.

1. Is it possible to explain phenomena of interference and diffraction on the supposition that there are only light-quanta without anything like the electromagnetic field of old theories? We might try to account for the observed phenomena by considering light as consisting in changes of state confined to very small spaces. The periodicity which characterizes a beam of light might be due to certain periodical changes going on in the quanta themselves; they might, so to speak, carry the vibrations along with them.

It is clear that a quantum in which there is an internal period T, when moving forward with the speed c, will cause equal phases to be regularly spaced along its path at intervals cT, which could be termed wave-lengths. On this view the phenomena of interference and diffraction which require a motion of the quanta in curved lines (compare question 3 below) would have to be ascribed to certain mutual actions between the quanta, determining their motion. This would obviously require a sufficiently great number of quanta, and one can therefore judge of the possibility of a "pure" quantum theory of light by attending to the number of quanta involved in actual phenomena. This will be best explained by an example.

An opaque screen with a circular opening O, 0.4 millimeter in diameter, was placed before an incandescent lamp (with frosted bulb). The rays passing through O were received at a distance of 48 centimeters on a diffracting screen A with two small openings (of course one opening might have done just as well) very near each other and having a total free area of 0.25 square millimeter. The diffraction produced by them was observed in a plane B, 16 centimeters behind A.

On account of the weakness of the light few details could be distinguished, but two bright spots, s_1 and s_2, each surrounded by a bright ring, and a bright interference band bisecting s_1s_2 at right angles could be seen clearly. This diffraction pattern made the impression that if we wished to reproduce it by a distribution of bright spots, some thirty or fifty of them would certainly be required.

From the energy consumption in the lamp, and from the distances and dimensions, reckoning that 6 per cent of the energy emitted is

in the form of light, and that all the rays have the same frequency (that of yellow light), so that the quanta are all of one kind, I estimated the number of quanta which the plane B receives per second at $7 \cdot 10^7$. As the impressions received by the eye are accumulated during about 0.1 second, the number of quanta is found to be more than sufficient for delineating on the plane B, or rather on the retina, the diffraction pattern observed.

A serious difficulty arises, however, when we consider the distribution of the quanta in space. We must not forget that in the space behind A the $7 \cdot 10^7$ quanta transmitted through its openings during a second are distributed over a length of $3 \cdot 10^{10}$ centimeters. This means that the distance at which they follow each other is about 400 centimeters, and that in the mean the region between the planes A and B (that is, the region where the quanta have to be guided in their course) contains no more than 0.04 quantum. By this the probability that at the same instant there are in that region a certain number, say n quanta, becomes $(0.04)^n$, which, for $n = 5$, for example, is very small. That two quanta should approach each other to within a micron is next to impossible, and so under the circumstances of the experiment by far the greater number of the quanta would travel from A to B without being acted on by any deflecting force.

In what precedes it has been tacitly assumed that the quanta have very small dimensions in all directions. If they were somewhat like thin fibers extending along the line of propagation, things would be different. But even so it is difficult to imagine an interaction. If we want, for example, a cross section of the beam between A and B to be traversed by twenty fibers (and this is not asking too much if by their interaction the details of the diffraction pattern are to be produced) the length of each fiber must be about 80 or 90 meters.

I think we may safely conclude from these considerations that there must be something else besides the quanta which governs their motion, and it is but natural to consider this as the function of the ordinary electromagnetic field, though perhaps, taken by itself, this field does not contain any appreciable energy and has to be enlivened by the quanta.

It need hardly be added that the conclusion to which we have been driven would have been strengthened if we had assumed, as of course we should have to do, that only quanta of nearly the same frequency can influence each other's motion.

2. Can light-quanta be made responsible for the energy-stress tensor of radiation? Modern physics has recognized in a beam of

light not only a certain density and flux of energy but also a momentum and a system of stresses,— quantities between all of which there is an intimate connection, which has been emphasized especially in the theory of relativity. Now a system composed of a great number of moving particles also has an energy-stress tensor (compare § 37 and § 38), and the question arises whether it is possible to account, by the assumption of light-quanta, not only for the energy, the flux of energy, and the momentum of a beam, but also for the stresses existing in it. This can really be done in very simple cases, such as that of a single beam of parallel rays, but as a rule it is impossible. The following example will illustrate this:

Let us consider the superposition of two beams of light, both as simple as can be and propagated in different directions. Let these directions lie in the plane XOY, making with OX the angles α and $-\alpha$. Then, if the electric forces are in the direction of OZ, one beam may be represented by

$$E_z = a \cos n\left(t - \frac{x \cos \alpha + y \sin \alpha}{c}\right), \quad H_x = \sin \alpha \cdot E_z,$$
$$H_y = -\cos \alpha \cdot E_z,$$

and the other by

$$E_z = a \cos n\left(t - \frac{x \cos \alpha - y \sin \alpha}{c}\right), \quad H_x = -\sin \alpha \cdot E_z,$$
$$H_y = -\cos \alpha \cdot E_z,$$

the remaining components being zero.

Compounding the two, we find

$$E_z = 2\, a \cos \frac{ny \sin \alpha}{c} \cos n\left(t - \frac{x \cos \alpha}{c}\right),$$
$$H_x = 2\, a \sin \alpha \sin \frac{ny \sin \alpha}{c} \sin n\left(t - \frac{x \cos \alpha}{c}\right),$$
$$H_y = -2\, a \cos \alpha \cos \frac{ny \sin \alpha}{c} \cos n\left(t - \frac{x \cos \alpha}{c}\right).$$

Let us deduce from this the normal stress X_x. This is given by

$$X_x = \tfrac{1}{2}(E_x^2 - E_y^2 - E_z^2) + \tfrac{1}{2}(H_x^2 - H_y^2 - H_z^2),$$

and we find

$$X_x = -2\, a^2 (1 + \cos^2 \alpha) \cos^2 \frac{ny \sin \alpha}{c} \cos^2 n\left(t - \frac{x \cos \alpha}{c}\right)$$
$$+ 2\, a^2 \sin^2 \alpha \sin^2 \frac{ny \sin \alpha}{c} \sin^2 n\left(t - \frac{x \cos \alpha}{c}\right),$$

or, averaging over a long time,

$$X_x = -a^2(1 + \cos^2 \alpha) \cos^2 \frac{ny \sin \alpha}{c} + a^2 \sin^2 \alpha \sin^2 \frac{ny \sin \alpha}{c}.$$

The integral of this expression with respect to y taken for an interval that contains many wave-lengths is negative (pressure of radiation), but in some parts of a plane perpendicular to OX the stress is positive, reaching the value $a^2 \sin^2 \alpha$ at the points where $\frac{ny \sin \alpha}{c}$ is an odd multiple of $\frac{1}{2}\pi$. Now this can never be realized by means of quanta, any more than a negative pressure can result from the motion of molecules. The transfer of momentum by moving particles always produces negative values of the normal stresses X_x, Y_y, Z_z; this is due to the fact that the velocity of a particle and its momentum always have the same direction.

We shall next examine the question how far the distribution of energy in the radiation field and the flux of energy can be ascribed to quanta. Here again the above example may serve us. According to the ordinary formulæ the energy per unit of volume is given by

$$U = 2\,a^2(1 + \cos^2 \alpha)u^2 + 2\,a^2 \sin^2 \alpha \cdot v^2, \tag{351}$$

and the flow of energy has the components

$$S_x = 4\,a^2 c \cos \alpha \cdot u^2, \quad S_y = 4\,a^2 c \sin \alpha \cdot uv. \tag{352}$$

The meaning of u and v in these expressions is

$$u = \cos \frac{ny \sin \alpha}{c} \cos n\left(t - \frac{x \cos \alpha}{c}\right),$$

$$v = \sin \frac{ny \sin \alpha}{c} \sin n\left(t - \frac{x \cos \alpha}{c}\right).$$

Now any values of U (positive, of course) and of the components of S can be reproduced by means of quanta each of which has the definite energy ϵ, if only the equation of continuity,

$$\frac{\partial S_x}{\partial x} + \frac{\partial S_y}{\partial y} + \frac{\partial U}{\partial t} = 0, \tag{353}$$

is satisfied, as it is in our case. If we impose no condition as to the velocity of the quanta, we may even assume that all those which at a definite instant are contained in an element of volume are moving in the same way, so that at a given time and place there is but one stream of quanta. Indeed, it will suffice to take for the number of quanta per unit of volume

$$\frac{U}{\epsilon},$$

and to determine their velocity $\mathbf{v}$ by the vector equation

$$\mathbf{v} = \frac{\mathbf{S}}{U}.$$

We shall, however, adhere to the supposition that in free space the quanta can have no other velocity but that of light. If so, the interpretation just given will be possible only when v turns out to be equal to c, and this requires

$$S_x^2 + S_y^2 = c^2U^2. \tag{354}$$

The values (351) and (352) will be found not to agree with this relation and the hypothesis of a single stream of quanta must therefore be discarded.

A satisfactory solution may, however, be found by assuming the existence of a double stream, two sets of quanta moving independently of each other. If the above values are decomposed into the parts

$$\left.\begin{aligned} U_1 &= a^2(1+\cos\alpha)^2u^2 + a^2\sin^2\alpha\cdot v^2 \\ S_{1x} &= a^2c(1+\cos\alpha)^2u^2 - a^2c\sin^2\alpha\cdot v^2 \\ S_{1y} &= 2\,a^2c(1+\cos\alpha)\sin\alpha\cdot uv \end{aligned}\right\}, \tag{355}$$

and

$$\left.\begin{aligned} U_2 &= a^2(1-\cos\alpha)^2u^2 + a^2\sin^2\alpha\cdot v^2 \\ S_{2x} &= -\,a^2c(1-\cos\alpha)^2u^2 + a^2c\sin^2\alpha\cdot v^2 \\ S_{2y} &= 2\,a^2c(1-\cos\alpha)\sin\alpha\cdot uv \end{aligned}\right\}, \tag{356}$$

each part taken by itself satisfies both conditions (353) and (354), and we can therefore ascribe to the two sets of quanta the densities

$$\frac{U_1}{\epsilon} \quad \text{and} \quad \frac{U_2}{\epsilon}$$

and determine the directions in which they move by the ratios

$$\frac{S_{1y}}{S_{1x}} \quad \text{and} \quad \frac{S_{2y}}{S_{2x}}.$$

The mean values of S_{1x} and S_{2x} taken over a long lapse of time and a wide range of values of y are

$$a^2c\cos^2\tfrac{1}{2}\,\alpha\cos\alpha, \quad \text{and} \quad a^2c\sin^2\tfrac{1}{2}\,\alpha\cos\alpha,$$

both positive so long as $\alpha < \frac{1}{2}\,\pi$. The mean values of S_{1y} and S_{2y} are zero, so that in this case the main direction of both currents, apart from all complexities, is in the direction of the positive x, as was to be expected.

It is interesting to note what becomes of the formulæ for $\alpha = 0$ or $\alpha = \frac{1}{2}\,\pi$. In the first case the second current vanishes and the first is in the direction of x,

$$U_1 = 4\,a^2\cos^2 n\left(t - \frac{x}{c}\right), \quad S_1 = cU_1,$$

agreeing with what we found for a single beam. In the second case we have standing waves perpendicular to OY with the values

$$U_1 = a^2(u^2+v^2),\quad S_{1x} = a^2c(u^2-v^2),\quad S_{1y} = 2\,a^2cuv,$$
$$U_2 = a^2(u^2+v^2),\quad S_{2x} = a^2c(v^2-u^2),\quad S_{2y} = 2\,a^2cuv,$$
$$u = \cos\frac{ny}{c}\cos nt,\quad v = \sin\frac{ny}{c}\sin nt.$$

It must be owned that in this case the decomposition to which we have been led because we wished the quanta to move with the velocity c looks rather artificial. We should not expect that in the standing waves now under consideration, in which all quantities are independent of the coördinate x, quanta would ever move in a direction not normal to the waves, and it must be remarked that we could meet just as well all conditions that have been introduced by attributing the values $a^2c(u^2-v^2)$ and $a^2c(v^2-u^2)$, not to S_{1x} and S_{2x} but to S_{1z} and S_{2z}. However, the mean values of these expressions for long ranges of y and t vanish.

Instead of trying to interpret by means of quanta all the rapid fluctuations in the energy and its flow, we could also be satisfied with a representation of the mean state of things as it is during a long lapse of time; this, of course, would require simply a steady flow of quanta. For instance, in a single beam of light we should have a density of energy $\frac{1}{2}a^2$ and a flow $\frac{1}{2}a^2c$, and these can easily be conceived to be due to quanta.

In the case considered in what precedes, the product uv will be zero on an average, and the mean values of U, S_x, and S_y become, according to equations (351) and (352),

$$\left.\begin{aligned} U &= a^2(1+\cos^2\alpha)\gamma^2 + a^2\sin^2\alpha\cdot\sigma^2 \\ S_x &= 2\,a^2c\cos\alpha\cdot\gamma^2,\quad S_y = 0 \end{aligned}\right\}, \tag{357}$$

where γ and σ are abbreviations for

$$\cos\frac{ny\sin\alpha}{c}\quad\text{and}\quad\sin\frac{ny\sin\alpha}{c}.$$

In attempting to decompose this into two parts, as we did with the values (351) and (352), it is natural to assume that the separate parts of U, S_x, and S_y, which now must be constants, are also, like U and S themselves, independent of x. This implies, by virtue of the equation of continuity, that S_{1y} and S_{2y} are independent of y. Let us therefore try the decomposition

$$U_1 = p_1\gamma^2 + q_1\sigma^2,\quad S_{1x} = c(r_1\gamma^2 + s\sigma^2),\quad S_{1y} = ck,$$
$$U_2 = p_2\gamma^2 + q_2\sigma^2,\quad S_{2x} = c(r_2\gamma^2 - s\sigma^2),\quad S_{2y} = -ck,$$

where p_1, p_2, q_1, q_2, r_1, r_2, s, and k are constants. The sums $p_1 + p_2$, $q_1 + q_2$, $r_1 + r_2$ are known by (357), and the application, to each of the two parts separately, of the condition (354), which must hold for any value of y, leads to the relations

$$r_1^2 + k^2 = p_1^2, \qquad r_2^2 + k^2 = p_2^2, \tag{358}$$

$$r_1 s + k^2 = p_1 q_1, \quad -r_2 s + k^2 = p_2 q_2, \tag{359}$$

$$s^2 + k^2 = q_1^2, \qquad s^2 + k^2 = q_2^2. \tag{360}$$

By the last two of these it appears that $q_2^2 = q_1^2$ and since $q_1 + q_2$ is different from 0, we must have $q_2 = q_1$, their common value being

$$q = \tfrac{1}{2}\, a^2 \sin \alpha.$$

On the other hand, p_1 and p_2 cannot be equal, for if they both had the value $p = \frac{1}{2}(p_1 + p_2) = \frac{1}{2}\, a^2(1 + \cos^2 \alpha)$, we should infer from (359) (because $r_1 + r_2$ is different from 0) $s = 0$; but then there is a contradiction between (359) and (360), because p is not equal to q.

If, however p_1 and p_2 are different, we can deduce from (358) and (359)

$$s^2 = q^2 - \frac{2\,q}{p_1 + p_2}\, k^2,$$

and this, when combined with (360), again leads us to the contradiction $p = q$. Our conclusion must therefore be that it is impossible to account for the time averages of U and S by two steady flows of quanta.

3. Can quanta move in curved lines? When, for a system of quanta, we know the values of U and S satisfying the conditions (353) and (354), we can determine not only the direction of motion at any time and place but also the paths of the individual quanta. In cases like those represented by (355) or (356) these paths will be curved lines whose radius of curvature is of the order of magnitude of the wave-length, so that the motion of each quantum is far from simple.

That the quanta cannot move in straight lines can also be seen by elementary considerations. Take, for example, the interference produced by two in all respects equal luminous points P and Q, somewhat like Fresnel's experiment with the two mirrors. The loci in space of the minima of intensity will be certain hyperboloids having their foci at P and Q. A straight line L must necessarily intersect surfaces of this kind; and if this takes place at a point R, a quantum moving along L would strike a screen placed at R at a point of a dark interference fringe.

Now if the quanta are to be guided in their motion by the electromagnetic field, and if we wish to maintain the principle of the

conservation of momentum, the curvature of the trajectories implies that a certain momentum can have its seat in the field. And here it is to be understood that this would not be the ordinary electromagnetic momentum that has been relegated to the quanta, but a momentum of some new kind. We should have to develop a new theory of momenta and stresses.

Of course it might well be that when we consider spaces of a certain extent or intervals of time of a certain length, there is compensation of the interchanges of momentum between the field and the quanta. When, however, as in the case examined in question 1, the region where the interference or diffraction takes place contains scarcely more than one quantum at a time, the momentum which it receives ought certainly to be taken from the field in that region.

Note 17, § 54

The proof of the theorem here given is briefly as follows:

1. Let the radiation consist of beams whose width is great in comparison with the wave-length, and let the two media be separated by a plane surface; the results obtained on this last supposition will remain valid for a curved boundary if its curvature is sufficiently small.

Each beam is laterally limited by a cylindrical surface whose generating lines may be called rays, and each incident beam will give rise to a certain number of reflected or refracted beams. For all beams belonging together in this way the part Ω of the boundary which they intercept is the same.

Understanding by τ a definite interval of time, we can intersect each beam by two planes in the direction of the wave-front, at a distance $w\tau$ from each other, where w is the group-velocity of the waves. The energy contained at a certain instant in the part of the incident beam thus delimited will be found at some later moment in similar parts of the reflected and refracted beams. We shall denote any of these spaces by S.

Let α and β be the sharp angles which the ray makes with the normal N to the boundary and the wave-normal respectively. Then the part of a wave-front lying within the beam is

$$\Omega \frac{\cos \alpha}{\cos \beta},$$

and the space S has the magnitude

$$S = \Omega w\tau \frac{\cos \alpha}{\cos \beta}. \tag{361}$$

2. The wave-normals of beams which belong together lie in a plane passing through the normal N to the boundary (plane of incidence), and if θ is the acute angle between N and the wave-normal, the ratio

$$\frac{v}{\sin\theta}$$

will be the same for all the beams in question. Here v is the velocity of the waves; that is, the speed of propagation of the phases.

If χ is the angle between the plane of incidence and a fixed plane passing through the normal N, the direction of a wave-normal can be determined by θ and χ. The velocity v will in general be a function of both these angles.

For two beams belonging together

$$\frac{v'}{\sin\theta'} = \frac{v}{\sin\theta}, \quad \chi' = \chi, \tag{362}$$

from which we can deduce the ratio between corresponding cones, $d\omega$ and $d\omega'$, for the two beams. If these cones were characterized by $d\theta$, $d\chi$ and $d\theta'$, $d\chi'$, respectively, their openings would be $\sin\theta\, d\theta\, d\chi$ and $\sin\theta' d\theta' d\chi'$. It can be inferred from this that when $d\omega$, $d\omega'$ belong together,

$$\frac{d\omega'}{d\omega} = \frac{\sin\theta'}{\sin\theta}\frac{d(\theta', \chi')}{d(\theta, \chi)},$$

where the last factor is the functional determinant of θ', χ' with respect to θ, χ (to be taken with the positive sign). Since, according to (362),

$$\frac{\partial\chi'}{\partial\theta} = 0, \quad \frac{\partial\chi'}{\partial\chi} = 1,$$

this becomes

$$\frac{d\omega'}{d\omega} = \frac{\sin\theta'}{\sin\theta}\frac{d\theta'}{d\theta}.$$

Here we have written $\frac{d\theta'}{d\theta}$ instead of $\frac{\partial\theta'}{\partial\theta}$ because in what follows we need consider only one position of the plane of incidence. In this plane v and v' are definite functions of θ and θ', and we find from (362)

$$\left(-\frac{\cos\theta'}{\sin^2\theta'}v' + \frac{1}{\sin\theta'}\frac{dv'}{d\theta'}\right)\frac{d\theta'}{d\theta} = -\frac{\cos\theta}{\sin^2\theta}v + \frac{1}{\sin\theta}\frac{dv}{d\theta};$$

thus

$$\frac{d\omega'}{d\omega} = \frac{v'^2}{v^2}\cdot\frac{v\cos\theta - \frac{dv}{d\theta}\sin\theta}{v'\cos\theta' - \frac{dv'}{d\theta'}\sin\theta'}. \tag{363}$$

3. In any of the cases to be considered the wave-surface and the direction of the ray corresponding to a given wave-normal are found by a well-known construction. If, from a fixed point O, a line OP is drawn whose length represents the velocity v of waves perpendicular to it, and if through P a plane V is made to pass at right angles to this line, the surface which envelops all the planes V, found by giving to OP different directions, is the wave-surface, and the line joining O to the point R where it is touched by the plane V is the ray corresponding to OP as wave-normal.

We shall now confine ourselves to directions of OP lying in a definite plane of incidence. Let OX and OY be two axes in this plane which are at right angles to each other, and the first of which, making the acute angle θ with OP, is parallel to the normal N to the boundary. Then the equation of a line L drawn through P at right angles to OP will be

$$x \cos \theta + y \sin \theta = v. \tag{364}$$

Changing θ by $d\theta$, we obtain a second line L' intersecting L at a point Q, and OQ will be the projection of OR on the plane of incidence. This is clear because planes passing through L and L' and perpendicular to the plane of incidence must both contain OR.

The coördinates x_Q and y_Q of the point Q satisfy equation (364) and also the relation, deduced from it by differentiation,

$$-x \sin \theta + y \cos \theta = \frac{dv}{d\theta},$$

giving

$$x_Q = v \cos \theta - \frac{dv}{d\theta} \sin \theta.$$

4. The cosines of the angles α and β which occur in (361) are proportional to the projections of OR on OX and OP respectively, that is, to the projections of OQ on those lines, because QR is perpendicular to the plane of incidence. But the projections of OQ are x_Q and v, so that

$$\frac{\cos \alpha}{\cos \beta} = \frac{1}{v}\left(v \cos \theta - \frac{dv}{d\theta} \sin \theta\right).$$

Combining this with (361) and (363), we are led to the conclusion that the expression

$$\frac{S\,d\omega}{v^2 w}$$

has the same value for all the beams that belong together.

5. When, in what follows, we shall have to speak of the amounts of energy contained in the parts $S_1, S_2, \cdots$ of the different beams,

we shall always have in view only those parts of the electromagnetic field for which the frequency lies in the interval $d\nu$ and the wave-normal within the cone $d\omega_1$, or $d\omega_2$, etc., as the case may be.

Now if B_1 is an incident beam, giving rise to the reflected or refracted beams $B_2, B_3, \cdots$, and if at a certain instant the space S_1 contains an amount of energy E_1, we shall find, after a certain time T, a fraction $k_{12}E_1$ of this energy in S_2, another fraction $k_{13}E_1$, in S_3, and so on, and we shall have

$$k_{12} + k_{13} + \cdots = 1. \tag{365}$$

We pass on from this to the case in which B_1 is a beam propagated away from the boundary. Different incident beams have contributed to its formation; namely, beams going along the paths $B_2, B_3, \cdots$, and no others. At a definite time t the spaces $S_1, S_2, S_3, \cdots$ will contain the quantities of energy $E_1, E_2, E_3, \cdots$. What will be the energy in S_1 at the time $t + T$?

At that instant the energy E_1 will have left the space considered, but its place will be occupied by certain fractions of the energies $E_2, E_3, \cdots$, say $k_{21}E_2, k_{31}E_3, \cdots$, so that the total energy in S_1 has become $k_{21}E_2 + k_{31}E_3 + \cdots$. The distribution of energy will therefore be stationary when, for any space S_1, we have

$$k_{21}E_2 + k_{31}E_3 + \cdots = E_1. \tag{366}$$

This condition is fulfilled when, with the restrictions specified by $d\omega$ and $d\nu$, the amount of energy per unit of volume is given by

$$\frac{A\, d\omega\, d\nu}{v^2 w},$$

where A is a function of the frequency and the temperature that is the same for all bodies; for in this case, on account of the conclusion to which we came just now, the quantities $E_1, E_2, \cdots$ are all equal, and equation (366) is satisfied because, by Helmholtz's proposition, $k_{21} = k_{12}$, $k_{31} = k_{13}, \cdots$, so that, by virtue of (365),

$$k_{21} + k_{31} + \cdots = 1.$$

We shall not give the proof of Helmholtz's theorem, but it may be remarked that it is not at all self-evident. In fact, the coefficients k_{12} and k_{21} refer to very different cases. In one the energy of an incident beam B_1 is partitioned between the paths $B_2, B_3, \cdots$, and in the other case the energy arriving along the path B_2 is divided between B_1 and other paths of which there was no question in the former partition.

Note 18, § 58

The condition
$$\delta \int ds = 0, \tag{367}$$

by means of which Einstein determines the motion of a particle in the gravitation field, may be put in the form of Hamilton's theorem in ordinary dynamics.

Let L be Lagrange's function expressed in terms of the coördinates x_1, x_2, x_3 and the velocity components $\dot{x}_1$, $\dot{x}_2$, $\dot{x}_3$ of a particle, and suppose that the parameters contained in L are given functions of the time; we shall denote by $\frac{\partial}{\partial t}$ a differentiation with respect to the time in so far as it appears in these parameters, so that we have

$$\frac{dL}{dt} = \sum \dot{x} \frac{\partial L}{\partial x} + \sum \ddot{x} \frac{\partial L}{\partial \dot{x}} + \frac{\partial L}{\partial t}. \tag{368}$$

In this equation, and in others that are to follow, the sums consist of three terms in which we must take for x the variables x_1, x_2, x_3 respectively.

In Hamilton's theory the real motion of the particle is compared with a varied motion which differs from it to an infinitely slight extent, the coördinates of the particle at a definite instant t being x_1, x_2, x_3 in the original motion and $x_1 + \delta x_1$, $x_2 + \delta x_2$, $x_3 + \delta x_3$ in the varied one. When the variations δx have been chosen as functions of the time, the varied motion with the velocities occurring in it is completely known. The symbol d will be used to indicate the changes that occur in an element of time dt, and the sign δ will refer to the virtual change of any quantity when, for a fixed value of t, we pass from the original motion to the modified one. This involves the relations

$$\delta \dot{x} = \frac{d\,\delta x}{dt}.$$

Let the particle be acted on by an external force with the components X_1, X_2, X_3. Then Hamilton's theorem is expressed by the equation

$$\delta \int_{t_1}^{t_2} L\,dt + \int_{t_1}^{t_2} \sum X\,\delta x\,dt = 0, \tag{369}$$

where it is to be understood that the initial position (at time t_1) and the final position (at t_2) are left unaltered, so that both for $t = t_1$ and $t = t_2$ we have $\delta x = 0$.

The equations of motion are easily deduced from (369). We have

$$\delta L = \sum \frac{\partial L}{\partial x}\,\delta x + \sum \frac{\partial L}{\partial \dot{x}}\,\delta \dot{x} = \sum \frac{\partial L}{\partial x}\,\delta x + \sum \frac{\partial L}{\partial \dot{x}}\,\frac{d\delta x}{dt}$$

$$= \sum \left\{ \frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) \right\} \delta x + \frac{d}{dt} \sum \frac{\partial L}{\partial \dot{x}}\,\delta x.$$

If this is substituted in the first term of (369), for which we may write $\int_{t_1}^{t_2} \delta L\,dt$, the last part of δL disappears; and since the variations δx may be arbitrarily chosen, the coefficient of each of them must be zero. Thus we have the three equations of motion

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = X. \qquad (370)$$

It is natural to interpret these formulæ by calling the quantities

$$\frac{\partial L}{\partial \dot{x}_1},\quad \frac{\partial L}{\partial \dot{x}_2},\quad \frac{\partial L}{\partial \dot{x}_3} \qquad (371)$$

the components of the momentum of the particle, and

$$\frac{\partial L}{\partial x_1},\quad \frac{\partial L}{\partial x_2},\quad \frac{\partial L}{\partial x_3}$$

those of the force acting on it, in so far as this force is contained in the function L.

We next consider the equation of energy. Multiplying the equations of motion by $\dot{x}_1$, $\dot{x}_2$, $\dot{x}_3$, and adding, we find, by a simple transformation and by using (368),

$$\frac{d}{dt}\left(\sum \dot{x}\,\frac{\partial L}{\partial \dot{x}} - L\right) + \frac{\partial L}{\partial t} = \sum X\dot{x}. \qquad (372)$$

We interpret this by defining

$$\dot{x}_1\,\frac{\partial L}{\partial \dot{x}_1} + \dot{x}_2\,\frac{\partial L}{\partial \dot{x}_2} + \dot{x}_3\,\frac{\partial L}{\partial \dot{x}_3} - L \qquad (373)$$

as the energy of the particle and by saying that its rate of change, in so far as it is due to L, is

$$-\frac{\partial L}{\partial t}.$$

These formulæ may be applied to the case of the gravitation field if we put

$$L = -m\,\frac{ds}{dt},$$

which makes (369) if there are no external forces, identical with Einstein's condition (367). We may write for this function L

$$L = -m\sqrt{\sum (ab)\, g_{ab}\dot{x}_a\dot{x}_b} \tag{374}$$

(where $\dot{x}_4 = 1$), and this is what we have to substitute in the expressions (371) for the momenta. Calling these $-q_1$, $-q_2$, $-q_3$, we find for $a = 1, 2, 3$

$$q_a = m\frac{\sum (b)\, g_{ab}\dot{x}_b}{\sqrt{\sum (ab)\, g_{ab}\dot{x}_a\dot{x}_b}},$$

where it has been taken into account that, for a chosen a,

$$\frac{\partial}{\partial \dot{x}_a}\sum (mn)\, g_{mn}\dot{x}_m\dot{x}_n = 2\sum (b)\, g_{ab}\dot{x}_b.$$

In a simpler form, $$q_a = \frac{m}{ds}\sum (b)\, g_{ab}dx_b. \tag{375}$$

This same formula holds for $a = 4$ if by q_4 we understand the energy. Indeed, since (374) is a homogeneous linear function of the variables $\dot{x}_a$, we have

$$\sum (a)\, \dot{x}_a \frac{\partial L}{\partial \dot{x}_a} = L,$$

so that the expression (373) for the energy becomes

$$-\frac{\partial L}{\partial \dot{x}_4}.$$

Therefore we have for $a = 4$, as well as for the other values,

$$q_a = -\frac{\partial L}{\partial \dot{x}_a}.$$

As to the equations of motion (370) and the equation of energy (372), dropping the terms with X, we can now write them in the form ($c = 1, 2, 3, 4$),

$$\frac{dq_c}{dt} = -\frac{\partial L}{\partial x_c} = \frac{1}{2}\frac{m}{ds\, dt}\sum (ab)\, \frac{\partial g_{ab}}{\partial x_c}\, dx_a dx_b, \tag{376}$$

which shows the meaning of the quantity ω in equation (259).

It is interesting to note that these four equations are found simultaneously when we determine the course in the space R_4 of a geodesic line between two given points. In doing this we shall vary the four coördinates x_a by δx_a. The sign δ will always refer to a point of the original line and the corresponding point of the varied

line, and the sign d to the extremities of a line-element. It is easily seen that $\delta dx_a = d\delta x_a$.

Starting from (367), or, rather, what amounts to the same, from

$$\int \frac{\delta(ds^2)}{ds} = 0, \tag{377}$$

the calculation is as follows:

$$\delta(ds^2) = \delta \sum (ab)\, g_{ab} dx_a dx_b = \sum (abc)\, \frac{\partial g_{ab}}{\partial x_c} dx_a dx_b \delta x_c$$
$$+ \sum (ab)\, g_{ab} dx_a \delta dx_b + \sum (ab)\, g_{ab} dx_b \delta dx_a,$$

or, if in the second term b is replaced by c, and in the last a by c and b by a,

$$\frac{\delta(ds^2)}{ds} = \frac{1}{ds} \sum (abc)\, \frac{\partial g_{ab}}{\partial x_c} dx_a dx_b \delta x_c + 2 \sum (ac)\, g_{ac} \frac{dx_a}{ds} d\delta x_c.$$

In the integral the last term may be replaced by

$$-2 \sum (ac)\, d\left(g_{ac} \frac{dx_a}{ds}\right) \delta x_c,$$

and we deduce from (377), if we equate to zero the coefficient of each δx_c and divide by dx_4,

$$\frac{1}{ds\, dt} \sum (ab)\, \frac{\partial g_{ab}}{\partial x_c} dx_a dx_b - 2 \frac{d}{dt} \sum (a)\, g_{ac} \frac{dx_a}{ds} = 0,$$

agreeing with (376).

Note 19, § 58

We shall now give a proof of the relation

$$\frac{N}{\sqrt{-g}\, dx_4} = \frac{N'}{\sqrt{-g'}\, dx'_4}.$$

Let Ω be a limited region in the four-dimensional space R_4, and let us consider the parts, lying within Ω, of the world-lines of the particles, and take the sum L of the lengths of all these parts, evaluating each length by the formula for ds. When Ω is infinitely small, and when the particles are so numerous that even in that case a great number of world-lines pass through Ω, the value of L will be proportional to Ω, and

$$Q = \frac{L}{\Omega} \tag{378}$$

may be called the sum of the lengths of world-lines per unit of space R_4.

This quantity will be independent of the shape of Ω. In order to find another expression for it, we define Ω as follows. We first

consider the points for which x_1, x_2, x_3 lie within a definite element dS of the three-dimensional space x_1, x_2, x_3, while x_4 has a definite value. Starting from these points, in so far as world-lines pass through them, we take on each world-line an element of a definite length ds in the direction in which the line is really described. Since the measurement of parts of R_4 is based on the convention that the magnitude of an element determined by $dx_1 \cdots dx_4$ shall be given by the product $dx_1 dx_2 dx_3 dx_4$, the volume of the element now defined will be

$$\Omega = dS\, dx_4,$$

where dx_4 is the change of x_4 when a point passes along ds.

On the other hand, the number of parts of world-lines is now $N\,dS$, and the length of each of them is ds, so that we have $L = N\,dS\,ds$ and

$$Q = N\frac{ds}{dx_4}. \tag{379}$$

All this may be said whatever be the system of coördinates. Thus, if we pass from x_a to x'_a, the two formulæ (378) and (379) will still hold though several of the quantities involved take new values.

In the first place, when, after the change, we apply (378) to the same region, the quantity L will remain the same because ds and the length of any line are invariant. But the numerical value of the magnitude of the element will become different, say Ω', and we shall have

$$Q' = \frac{L}{\Omega'}.$$

The ratio between Ω and Ω' is given by the functional determinant of the variables x_a with respect to x'_a; that is, if we use (230), by the determinant p formed with the elements p_{ab}, or rather by its absolute value $|p|$. Thus

$$\Omega = |p|\Omega',$$

and, consequently,

$$Q' = |p|Q.$$

On the other hand, we have, corresponding to (379),

$$Q' = N'\frac{ds}{dx'_4},$$

where we can take for ds the same element of the world-line as in (379), so that dx_4 and dx'_4 are the simultaneous changes of x_4 and x'_4 when we pass along the world-line. Combining these results, we find

$$\frac{N'}{dx'_4} = |p|\frac{N}{dx_4}. \tag{380}$$

It is shown in the theory of determinants that when the quantities g_{ab} and g'_{ab} are related to each other in the way expressed by (245), the determinants g' and g, formed the one with the quantities g'_{ab} and the other with g_{ab}, are to each other in the ratio

$$\frac{g'}{g} = p^2.$$

Therefore
$$|p| = \frac{\sqrt{-g'}}{\sqrt{-g}},$$

and (380) becomes
$$\frac{N'}{\sqrt{-g'} \cdot dx'_4} = \frac{N}{\sqrt{-g} \cdot dx_4}.$$

We may now write, by virtue of (269),

$$\frac{T_a^b}{\sqrt{-g}} = \frac{N}{\sqrt{-g} \cdot dx_4} q_a dx_b$$

and (270) is found directly, if one takes into consideration that the first factor is a scalar, dx_b a contravariant vector, and q_a, as is shown by (375), a covariant vector.

Note 20, § 58

We have
$$T_{ab} = \sum (c) \frac{1}{\sqrt{-g}} g_{bc} T_a^c,$$

$$T_a^c = N q_a v_c = \frac{N}{dx_4} q_a dx_c,$$

$$q_a = \frac{m}{ds} \sum (e) g_{ae} dx_e.$$

Thus
$$T_{ab} = \frac{Nm}{ds\, dx_4 \sqrt{-g}} \sum (ce) g_{bc} g_{ae} dx_c dx_e.$$

Here we may interchange the suffixes c and e because each of them is to be given the values 1, 2, 3, 4. If, after that, we interchange a and b, we are led back to the first form. Consequently

$$T_{ba} = T_{ab}.$$

Note 21, § 59

The object of this note is to give some further mathematical developments leading to the expression, in terms of the potentials and their derivatives, of the tensor that has been represented by H_{ab}^a.

1. Let ABC be a triangle in R_4 formed of three line-elements. Let AB and AC, taken in the direction of the succeeding letters, be

denoted by ds and $d's$, their components, that is, the differentials of the coördinates corresponding to them, by dx_a and $d'x_a$. Those of BC, which we shall call $d''s$, obviously are $d'x_a - dx_a$.

In the following equations we shall be concerned with quantities that are of the second order with respect to dx_a, $d'x_a$. As quantities of a higher order may be omitted, we may put, understanding by g_{ab} the values at A,

$$d''s^2 = \sum_{(ab)} g_{ab}(d'x_a - dx_a)(d'x_b - dx_b).$$

Also $\quad ds^2 = \sum_{(ab)} g_{ab} dx_a dx_b, \quad d's^2 = \sum_{(ab)} g_{ab} d'x_a d'x_b.$

The triangle will be said to be rectangular at the point A when

$$d''s^2 = ds^2 + d's^2.$$

If the above values are substituted in this equation, the condition required in order that ds and $d's$ may be at right angles to each other is found to be

$$\sum_{(ab)} g_{ab} dx_a d'x_b = 0. \tag{381}$$

2. The direction of a line-element may be specified by what we may term the "direction constants"

$$\xi^a = \frac{dx_a}{ds},$$

four quantities connected by the relation

$$\sum_{(ab)} g_{ab} \xi^a \xi^b = 1 \tag{382}$$

that follows directly from the formula for ds^2.

Since ds is a scalar and dx_a a contravariant vector, ξ^a is a vector of the same kind.

Dividing (381) by $ds\, d's$, we find the condition for two directions perpendicular to each other in the form

$$\sum_{(ab)} g_{ab} \xi^a \xi'^b = 0.$$

3. When two line elements ds and $d's$ beginning at the same point P are compounded (by addition of their components) into a third line element $d''s$, also drawn from P, the three are said to lie in a plane. If their direction constants are ξ^a, ξ'^a, and ξ''^a, respectively, we have

$$\xi''^a = \frac{ds}{d''s}\xi^a + \frac{d's}{d''s}\xi'^a. \tag{383}$$

We shall have to consider the special case that $d's$ is perpendicular to ds and infinitely small with respect to it. The first condition leads to $d''s^2 = ds^2 + d's^2$, and by virtue of the second (383) takes the form

$$\xi''^a = \xi^a + \epsilon\xi'^a,$$

where ϵ is infinitely small. It is easily seen that these values satisfy equation (382).*

4. The determination of the differential equation of a geodesic line has already been discussed in Note 18.

Let now the last division in that note be made, not by dx_4 but rather by ds, and introduce the direction constants. The result is

$$\sum_{(ab)} \frac{\partial g_{ab}}{\partial x_c} \xi^a \xi^b - 2\sum_{(a)} \left(\frac{dg_{ac}}{ds} \xi^a + g_{ac} \frac{d\xi^a}{ds} \right) = 0.$$

These are four equations ($c = 1, 2, 3, 4$) which enable us to determine the values of $d\xi^a/ds$ so that we shall find the law according to which the direction of a geodesic changes from point to point.

As the potential g_{ac} is a function of the coördinates, we can substitute

$$\frac{dg_{ac}}{ds} = \sum_{(k)} \frac{\partial g_{ac}}{\partial x_k} \xi^k.$$

By this we find

$$2\sum_{(a)} g_{ac} \frac{d\xi^a}{ds} = \sum_{(ab)} \frac{\partial g_{ab}}{\partial x_c} \xi^a \xi^b - 2\sum_{(ak)} \frac{\partial g_{ac}}{\partial x_k} \xi^a \xi^k.$$

This becomes simpler if in one half of the last term we replace k by b, and in the other half k by a and a by b, as follows:

$$2\sum_{(a)} g_{ac} \frac{d\xi^a}{ds} = \sum_{(ab)} \left(\frac{\partial g_{ab}}{\partial x_c} - \frac{\partial g_{ac}}{\partial x_b} - \frac{\partial g_{bc}}{\partial x_a} \right) \xi^a \xi^b.$$

Using the generally used notation

$$\begin{bmatrix} ab \\ c \end{bmatrix} = \frac{1}{2} \left(\frac{\partial g_{ac}}{\partial x_b} + \frac{\partial g_{bc}}{\partial x_a} - \frac{\partial g_{ab}}{\partial x_c} \right), \tag{384}$$

we are led to
$$\sum_{(a)} g_{ac} \frac{d\xi^a}{ds} = -\sum_{(ab)} \begin{bmatrix} ab \\ c \end{bmatrix} \xi^a \xi^b. \tag{385}$$

It remains to solve from these formulæ the four quantities $d\xi^a/ds$ which we wish to know. For this purpose we may profitably avail

* We may remark here that some of the quantities occurring in our formulæ may eventually have imaginary values. The coördinates with their differentials and the gravitation potentials will always be real, but the values that we shall give to the potentials are such that two classes of lines must be distinguished. For some lines ds^2 is positive (world-lines of particles), for others it is negative, the transition from one case to the other being made by the world-lines of light-signals which are characterized by $ds^2 = 0$. For a line of the second class ds becomes imaginary, and so do the direction constants. If in the case just considered ds is of the first and $d's$ of the second class, ϵ must be given an imaginary value so that ξ'^a, like ξ^a, may be real.

In these considerations, the aim of which is solely to give a general idea of the geometry in R_4, we need not go into the details of these questions. Suffice it to say that all the final equations (for example, the formulæ for H^a_{bkl} and G_{ab}, and the field equations) contain real quantities only.

ourselves of the quantities g^{ab} mentioned already in § 58, the set of which may be called the inverse to the set g_{ab}. The theorem that

$$\sum\nolimits_{(c)} g_{ac}g^{bc}$$

has the value 1 when $b = a$, and is zero when b differs from a, is often very useful. As is indicated already by the notation, these new quantities form a contravariant tensor of the second rank, and there is the equality $g^{ab} = g^{ba}$.

The solution of the four equations (385) is obtained by multiplying each of them by g^{ck}, where k is a definite number, and then adding them. The result is

$$\frac{d\xi^k}{ds} = -\sum\nolimits_{(ab)} \begin{Bmatrix} ab \\ k \end{Bmatrix} \xi^a \xi^b, \tag{386}$$

where

$$\begin{Bmatrix} ab \\ k \end{Bmatrix} = \sum\nolimits_{(c)} g^{ck} \begin{bmatrix} ab \\ c \end{bmatrix}, \tag{387}$$

a symbol representing a definite function of the potentials g_{ab} and their derivatives.

The expressions $\begin{bmatrix} ab \\ c \end{bmatrix}$ and $\begin{Bmatrix} ab \\ c \end{Bmatrix}$ are called "Christoffel's symbols." They have the properties expressed by the formulæ

$$\begin{bmatrix} ba \\ c \end{bmatrix} = \begin{bmatrix} ab \\ c \end{bmatrix} \quad \text{and} \quad \begin{Bmatrix} ba \\ c \end{Bmatrix} = \begin{Bmatrix} ab \\ c \end{Bmatrix},$$

the former of which is a direct consequence of the definition (384), whereas the latter follows from it by virtue of (387).

From this equation we easily infer

$$\begin{bmatrix} ab \\ k \end{bmatrix} = \sum\nolimits_{(c)} g_{ck} \begin{Bmatrix} ab \\ c \end{Bmatrix}.$$

5. We proceed to seek the formula for the parallel displacement that has been defined in § 59. Let L be a geodesic line drawn from a point P, and Q a second point of the line near P. The distance s from P to Q, measured along the line, will be considered as infinitely small, but for our purpose it is necessary in the next formulæ to retain terms with s^2.

If ξ^a are the direction constants of L, and if, unless it be otherwise indicated, we give all quantities which depend on the coördinates the values they have at P, we have, by Maclaurin's theorem,

$$x_{aQ} = x_{aP} + s\xi^a + \frac{1}{2}s^2 \frac{d\xi^a}{ds},$$

or, by virtue of (386),

$$x_{aQ} = x_{aP} + s\xi^a - \frac{1}{2}s^2 \sum\nolimits_{(kl)} \begin{Bmatrix} kl \\ a \end{Bmatrix} \xi^k \xi^l. \tag{388}$$

From the point P we draw a line-element PP' with the direction constants η^a, at right angles to PL; this is the element which we wish to displace parallel to itself along PL. For this purpose we draw from P a second geodesic line PL', infinitely near PL and having its first element in a plane with PP' and the first element of PL. For the direction constants of this new line at the point P we may write $\xi^a + \epsilon\eta^a$, where ϵ is infinitely small; only its first power will have to be retained.

If on PL' we take a point Q' at a distance s equal to the part PQ of the first line, the line joining Q and Q' will give us the direction of PP' displaced toward the point Q.

We find the coördinates of Q' if in (388) we give ξ^a the increment $\epsilon\eta^a$. Thus the components of the line QQ', say of $\bar{d}s$, will be

$$\bar{d}x_a = x_{aQ'} - x_{aQ} = \epsilon s\left[\eta^a - s\sum{}_{(kl)}\begin{Bmatrix}kl\\a\end{Bmatrix}\xi^k\eta^l\right], \tag{389}$$

where the last term is made up of two parts, the equality of which is seen by interchanging k and l.

Our object is to find the direction constants of QQ'. For this purpose we must divide (389) by $\bar{d}s$. As we wish to know the direction constants accurately to terms of order s, we must retain in $\bar{d}s$, as we did in $\bar{d}x_a$, terms with ϵ and with ϵ^2. In

$$\bar{d}s^2 = \sum{}_{(ab)}\, g_{ab\,(Q)}\,\bar{d}x_a\bar{d}x_b$$

we must therefore calculate the terms with ϵ^2 and ϵ^3, which means that in $g_{ab(Q)}$ we must go to terms of order ϵ.

As the coördinates of Q are

$$x_r + s\xi^r,$$

we must substitute

$$g_{ab(Q)} = g_{ab} + s\sum{}_{(r)}\frac{\partial g_{ab}}{\partial x_r}\,\xi^r.$$

From (389) we find

$$\bar{d}x_a\bar{d}x_b = \epsilon^2 s^2\left[\eta^a\eta^b - s\eta^a\sum{}_{(kl)}\begin{Bmatrix}kl\\b\end{Bmatrix}\xi^k\eta^l - s\eta^b\sum{}_{(kl)}\begin{Bmatrix}kl\\a\end{Bmatrix}\xi^k\eta^l\right].$$

Hence $\bar{d}s^2$ consists of four parts; namely,

(1) $\epsilon^2 s^2\sum{}_{(ab)}\,g_{ab}\eta^a\eta^b = \epsilon^2 s^2$,

(2) $\epsilon^2 s^3\sum{}_{(abr)}\dfrac{\partial g_{ab}}{\partial x_r}\,\xi^r\eta^a\eta^b$,

(3) $-\,\epsilon^2 s^3\sum{}_{(abkl)}\,g_{ab}\begin{Bmatrix}kl\\b\end{Bmatrix}\xi^k\eta^a\eta^l = -\,\epsilon^2 s^3\sum{}_{(akl)}\begin{bmatrix}kl\\a\end{bmatrix}\xi^k\eta^a\eta^l$,

and (4) a part that is equal to the third because it differs from it only by the interchange of the suffixes a and b.

After having replaced in part (2) b by l and r by k, and substituted the value

$$\frac{\partial g_{al}}{\partial x_k} = \begin{bmatrix} kl \\ a \end{bmatrix} + \begin{bmatrix} ak \\ l \end{bmatrix},$$

we find, collecting these results

$$\bar{d}s^2 = \epsilon^2 s^2 \left[1 + s \sum (akl) \left\{ \begin{bmatrix} ak \\ l \end{bmatrix} - \begin{bmatrix} kl \\ a \end{bmatrix} \right\} \xi^k \eta^a \eta^l \right] = \epsilon^2 s^2,$$

because the sums with $\begin{bmatrix} ak \\ l \end{bmatrix}$ and $\begin{bmatrix} kl \\ a \end{bmatrix}$ cancel each other, one taking the form of the other by an interchange of a and l. Thus we shall find the direction constants of QQ' when (389) is divided by ϵs. It appears, finally, that the changes in the direction constants of PP' accompanying the displacement from P to Q are given by

$$d\eta^a = -\sum (kl) \begin{Bmatrix} kl \\ a \end{Bmatrix} \eta^l dx_k, \tag{390}$$

where the change of x_k, that is, $\xi^k s$, has been represented by dx_k.

It is easily seen that (390) also applies to the displacement of a line-element constantly directed along the geodesic line; for if we write ξ instead of η, we obtain an equation equivalent to (386).

It is further evident that the element PP' which is normal to the geodesic line in P remains at right angles to it. Indeed, in the expression

$$\sum (ab)\, g_{ab} \xi^a \eta^b, \tag{391}$$

which is zero at the point P, g_{ab} changes by $\sum (k) \dfrac{\partial g_{ab}}{\partial x_k} dx_k,$

$$\xi^a \text{ by } -\sum (kl) \begin{Bmatrix} kl \\ a \end{Bmatrix} \xi^l dx_k, \text{ and } \eta^b \text{ by } -\sum (kl) \begin{Bmatrix} kl \\ b \end{Bmatrix} \eta^l dx_k,$$

and we shall find without difficulty that this leaves (391) constant.

It remains to consider the parallel displacement of an element having an arbitrarily chosen direction. The way in which this is to be done was explained in the text, and a simple reasoning will show that in this case also the above formula for $d\eta^a$ applies. Moreover, the equation holds for any line-element dx_k, because a given element can always be considered as the first of a geodesic line.

Our conclusion is, therefore, that the changes of the direction constants of an element, for which we shall now write ξ, caused by the parallel displacement along an element dx_k, are always given by

$$d\xi^a = -\sum (kl) \begin{Bmatrix} kl \\ a \end{Bmatrix} \xi^l dx_k. \tag{392}$$

6. Riemann's tensor enters into the expression which is obtained when we proceed to calculate the changes $\Delta\xi^a$ that take place when an element whose direction constants are ξ^a is displaced parallel to itself so that its first extremity moves round a closed line in R_4. We shall suppose the dimensions of the line to be infinitely small, and we shall confine ourselves to quantities of the second order with respect to these dimensions.

Let P be the starting-point, and write for any other point Q of the circuit

$$x_{aQ} - x_{aP} = x_a.$$

Then we have to calculate

$$\Delta\xi^a = -\sum{}^{(kl)} \int \begin{Bmatrix} kl \\ a \end{Bmatrix} \xi^l dx_k. \qquad (393)$$

Here we can replace the first factor by

$$\begin{Bmatrix} kl \\ a \end{Bmatrix} + \sum{}^{(c)} \frac{\partial}{\partial x_c} \begin{Bmatrix} kl \\ a \end{Bmatrix} x_c \qquad (394)$$

if for $\begin{Bmatrix} kl \\ a \end{Bmatrix}$ as well as for its derivative we take the value at P. Moreover, we must keep in mind that in (393) ξ^l means the direction constant such as it has become after the displacement from P to the point Q. Consequently, if by ξ^l also we wish to understand its value at P, we must replace ξ^l in (393) by the sum of ξ^l and the change which this quantity has undergone during the displacement; that is, according to (392), by

$$\xi^l - \sum{}^{(fh)} \begin{Bmatrix} fh \\ l \end{Bmatrix} \xi^f x_h, \qquad (395)$$

because we may here confine ourselves to the first order in x_h.

In the expressions (394) and (395) the quantities x_c and x_h are the only ones that change as we pass along the circuit; all the remaining ones are constant. Moreover, in the product the terms with $x_c x_h$ may be neglected and the term with neither x_c nor x_h will contribute nothing to the integral in (393) because taken over a closed circuit $\int dx_k = 0$. Thus the value of $\Delta\xi^a$ consists of two parts, one containing derivatives of Christoffel's expressions, and the other, products of two such expressions.

The first part is $-\sum{}^{(ckl)} \frac{\partial}{\partial x_c} \begin{Bmatrix} kl \\ a \end{Bmatrix} \xi^l \int x_c dx_k,$

and the second may be written

$$\sum{}^{(cklu)} \begin{Bmatrix} ku \\ a \end{Bmatrix} \begin{Bmatrix} cl \\ u \end{Bmatrix} \xi^l \int x_c dx_k$$

if in it we replace h, f, l by c, l, u.

It may now be remarked that when we have an expression of the form

$$M = \sum_{(ck)} N_{(ck)} \int x_c dx_k,$$

where N contains the two suffixes c and k, we may also write

$$M = \sum_{(ck)} N_{(kc)} \int x_k dx_c$$

or

$$M = -\sum_{(ck)} N_{(kc)} \int x_c dx_k,$$

because for a closed circuit

$$\int x_c dx_k + \int x_k dx_c = \int d(x_c x_k) = 0.$$

We may also write, taking half the sum of the first and the third expression for M,

$$M = \tfrac{1}{2} \sum_{(ck)} [N_{(ck)} - N_{(kc)}] \int x_c dx_k.$$

Applying this to the two parts of $\Delta\xi^a$, we find

$$\Delta\xi^a = \tfrac{1}{2} \sum_{(ckl)} H^a_{lck} \xi^l \int x_c dx_k, \tag{396}$$

where

$$H^a_{lck} = \frac{\partial}{\partial x_k}\left\{\begin{matrix} cl \\ a \end{matrix}\right\} - \frac{\partial}{\partial x_c}\left\{\begin{matrix} kl \\ a \end{matrix}\right\} + \sum_{(u)} \left[\left\{\begin{matrix} ku \\ a \end{matrix}\right\}\left\{\begin{matrix} cl \\ u \end{matrix}\right\} - \left\{\begin{matrix} cu \\ a \end{matrix}\right\}\left\{\begin{matrix} kl \\ u \end{matrix}\right\}\right] \tag{397}$$

By means of these expressions, taken with all possible suffixes, the change in direction of any element when it is displaced parallel to itself along any closed circuit can be calculated, and so far the quantities

$$H^a_{lck}$$

measure what may be called the curvature of R_4.

It ought to be noted that the meaning and in general the value of H^a_{lck} is altered when the suffixes are interchanged. A direct consequence of (397) is that

$$H^a_{lkc} = -H^a_{lck},$$

which implies that the expression is zero for $k = c$. By this the number of different quantities contained in the symbol (if two differing only in sign are not distinguished) reduces to 96.

7. It remains to prove the theorem (in agreement with which the notation has already been chosen) that H^a_{lck} is a tensor of the fourth rank, covariant with respect to the suffixes l, c, k and contravariant with respect to a. For this purpose we have to examine the relations between H^a_{lck} and the corresponding quantities H'^b_{mnq} to which one is led when the coördinates are changed from x_a to x'_a. Now, choosing in R_4 a definite closed circuit and a definite line-element that is displaced along it, both the same whether we use x_a or x'_a, we can

consider the changes $\Delta\xi'^a$ as well as the changes $\Delta\xi^a$. We shall try to deduce the former from the latter.

The quantities ξ^a and ξ'^a are connected with each other by the transformation formulæ for contravariant vectors, and these formulæ also apply to $\Delta\xi^a$ and $\Delta\xi'^a$, because $\Delta\xi^a$ is the difference between two values of ξ^a occurring at the same point, and $\Delta\xi'^a$ the difference between corresponding quantities ξ'^a. Moreover, on the right-hand side of equation (396) ξ^l, x_c, and dx_k may be expressed in terms of the corresponding quantities with primes. Here again we have to use the transformation formulæ for contravariant vectors. This is immediately seen in the case of ξ^l and dx_k, but it applies equally to x_c because this symbol represents the difference between the coördinates of points that are at an infinitely short distance from each other. Moreover, for the coefficients p_{ab} that occur in the transformation formulæ we may always take the values which they have at the definite point P of the circuit; if we took into consideration the small changes of p_{ab} as we follow the closed line, we should be led to quantities of a higher order than we need consider.

After all this the proof is easy, namely,

$$\Delta\xi'^b = \sum\nolimits_{(a)} \pi_{ab}\Delta\xi^a = \tfrac{1}{2}\sum\nolimits_{(ackl)} \pi_{ab}H^a_{lck}\xi^l\int x_c dx_k$$

$$= \tfrac{1}{2}\sum\nolimits_{(acklmnq)} \pi_{ab}p_{lm}p_{cn}p_{kq}H^a_{lck}\xi'^m\int x'_n dx'_q,$$

and for this we may write

$$\Delta\xi'^b = \tfrac{1}{2}\sum\nolimits_{(mnq)}H'^b_{mnq}\xi'^m\int x'_n dx'_q$$

if we put
$$H'^b_{mnq} = \sum\nolimits_{(ackl)} \pi_{ab}p_{lm}p_{cn}p_{kq}H^a_{lck}.$$

Of these two formulæ the first shows, when compared with (396), that in the new system of coördinates H'^b_{mnq} plays the same part as H^a_{lck} in the old one, so that H'^a_{lck} will be obtained from (397) when primes are added to all the symbols; the second expresses the proposition that was to be proved.

Note 22, § 59

That G_{ab} is a covariant tensor is proved by the following transformations, each step of which will be easily understood when we remember that

$$\sum\nolimits_{(c)} p_{fc}\pi_{kc} = 1, \text{ for } k = f \text{ and } 0 \text{ for } k \neq f,$$

$$G'_{ab} = \sum\nolimits_{(c)} H'^c_{abc} = \sum\nolimits_{(cdefk)} p_{da}p_{eb}p_{fc}\pi_{kc}H^k_{def}$$

$$= \sum\nolimits_{(def)} p_{da}p_{eb}H^f_{def} = \sum\nolimits_{(de)} p_{da}p_{eb}G_{de}.$$

The proof that G_{ab} is symmetrical is somewhat less simple. One sees from (397) that

$$G_{lc} = \sum (a) H^a_{lca}$$

consists of two parts, one containing derivatives of Christoffel's expressions, and the second, products of such expressions. By working out the sums, taking into account that a permutation of the suffixes a and u does not change a sum $\Sigma(au)$, it is seen directly that the second part of G_{lc} is not altered by interchanging l and c.

The first part of G_{lc} is

$$\sum (a) \frac{\partial}{\partial x_a} \begin{Bmatrix} cl \\ a \end{Bmatrix} - \sum (a) \frac{\partial}{\partial x_c} \begin{Bmatrix} al \\ a \end{Bmatrix},$$

and here it is at once clear that the first term has the property in question, so that we have only to consider the expression

$$\sum (a) \frac{\partial}{\partial x_c} \begin{Bmatrix} al \\ a \end{Bmatrix}.$$

We may write for it

$$\frac{\partial}{\partial x_c} \sum (ab)\, g^{ab} \begin{bmatrix} al \\ b \end{bmatrix} = \frac{1}{2} \frac{\partial}{\partial x_c} \sum (ab)\, g^{ab} \left(\begin{bmatrix} al \\ b \end{bmatrix} + \begin{bmatrix} bl \\ a \end{bmatrix} \right)$$
$$= \frac{1}{2} \frac{\partial}{\partial x_c} \sum (ab)\, g^{ab} \frac{\partial g_{ab}}{\partial x_l}.$$

Now the quantities g^{ab} are such that $x_a = \Sigma(b)\, g^{ab} \xi_b$ is the inversion of $\xi_a = \Sigma(b)\, g_{ab} x_b$, from which we can derive the following rule for finding g^{ab} when the potentials g_{ab} are given. First form the determinant g of the potentials and take its minor with respect to a particular g_{ab} (g_{ba} being considered as distinct from g_{ab}). Then the corresponding g^{ab} is found when that minor is divided by g. Hence, as the minor belonging to g_{ab} is $\frac{\partial g}{\partial g_{ab}}$, we have

$$g^{ab} = \frac{1}{g} \frac{\partial g}{\partial g_{ab}} = \frac{\partial \log g}{\partial g_{ab}},$$

giving

$$\frac{\partial}{\partial x_c} \sum (ab)\, g^{ab} \frac{\partial g_{ab}}{\partial x_l} = \frac{\partial}{\partial x_c} \sum (ab) \frac{\partial \log g}{\partial g_{ab}} \frac{\partial g_{ab}}{\partial x_l} = \frac{\partial^2 \log g}{\partial x_c\, \partial x_l},$$

and now it is seen that the value is not changed by a permutation of c and l.

Note 23, § 60

When we have to consider the gravitational field produced by a material system that is symmetrical around the center O and is at rest with respect to the axes of coördinates, we may remark in the

first place that the line-element ds in R_4 is compounded of a line-element in the ordinary three-dimensional space R_3 (which henceforth we shall call $d\sigma$) and an element dx_4 or dt. The element $d\sigma$ can be decomposed into one in the direction of r and another at right angles to it. If we call these $d_1\sigma$ and $d_2\sigma$, the expression for ds^2 may be taken to have the form

$$ds^2 = A\, d_1\sigma^2 + B\, d_2\sigma^2 + C\, dx_4^2,$$

the coefficients being functions of r.

Now, not confining ourselves to a plane passing through O, we may determine the position of a point P in the usual way by polar coördinates r, ϕ, ψ, the first being the distance from O, the second the angle between OP and OX, and the third the angle which the plane POX makes with a fixed plane passing through OX. We then have

$$d_1\sigma = dr, \quad d_2\sigma^2 = r^2(d\phi^2 + \sin^2\phi\, d\psi^2),$$

and, writing x_1, x_2, x_3 for r, ϕ, and ψ, we may take for the line-element in R_4

$$ds^2 = -u\, dx_1^2 - v(dx_2^2 + \sin^2 x_2\, dx_3^2) + w\, dx_4^2.$$

We shall now try to find out how far the coefficients u, v, w can be determined as functions of r by means of the field equations.

If the values of the potentials,

$$g_{11} = -u, \quad g_{22} = -v, \quad g_{33} = -v\sin^2 x_2, \quad g_{44} = w, \tag{398}$$

are substituted in the formula for G_{ab}, and if primes are used to indicate differentiation with respect to r, we find

$$\left.\begin{aligned} G_{11} &= -\frac{u'v'}{2\,uv} - \frac{v'^2}{2\,v^2} + \frac{v''}{v} - \frac{u'w'}{4\,uw} - \frac{w'^2}{4\,w^2} + \frac{w''}{2\,w} \\ G_{22} &= -1 - \frac{u'v'}{4\,u^2} + \frac{v''}{2\,u} + \frac{v'w'}{4\,uw} \\ G_{33} &= \sin^2 x_2 \cdot G_{22} \\ G_{44} &= \frac{u'w'}{4\,u^2} - \frac{v'w'}{2\,uv} + \frac{w'^2}{4\,uw} - \frac{w''}{2\,u} \end{aligned}\right\}, \tag{399}$$

the remaining G_{ab}'s all being zero.

In the space outside the attracting body the components T_{ab} are all zero, so that the field equations require $G_{ab} = 0$, which reduces to

$$G_{11} = 0, \quad G_{22} = 0, \quad G_{44} = 0. \tag{400}$$

It is important to notice that these equations are interdependent, because between the above expressions for G_{11}, G_{22}, G_{44} there is the relation

$$-\frac{1}{2}G'_{11}+\frac{u}{v}G'_{22}-\frac{u}{2w}G'_{44}=\left(-\frac{u'}{2u}+\frac{v'}{v}+\frac{w'}{2w}\right)G_{11}.$$

The consequence of this is that the functions u, v, w are not completely determined by the equations. Indeed, these are satisfied by

$$u=\frac{ab}{(b-w)^4}\frac{w'^2}{w},\quad v=\frac{a}{(b-w)^2},\tag{401}$$

where w is any function of r, and a and b are constants.

This solution is found when the equations $G_{44}=0$ and $G_{11}+\frac{u}{w}G_{44}=0$ are divided by $-\frac{w'}{4u}$ and $\frac{v'}{2v}$ respectively. The resulting equations admit of integration, giving

$$\frac{v^2w'^2}{uw}=\text{const.}\quad\text{and}\quad\frac{v'^2}{uvw}=\text{const.}$$

Hence $\frac{v'^2}{v^3w'^2}$ also must be constant, or, if v is considered as a function of w, $\frac{1}{v^3}\left(\frac{dv}{dw}\right)^2=\text{const.}$, from which the above value of v is easily derived. It then follows immediately that u is equal to

$$\frac{w'^2}{(b-w)^4w}$$

multiplied by a constant, and substitution in $G_{22}=0$ shows that this constant must have the value ab.

That the solution should contain an arbitrary constant such as b, which will depend on the mass of the attracting body, was to be expected; such a factor also occurs in the solutions of Laplace's equation. But it is a peculiar feature of Einstein's theory that even the form of the functions remains to a certain extent indeterminate. It has already been mentioned in the text that this is unavoidable in a theory whose fundamental principle is that the general equations have the same form in all systems of coördinates.

The indeterminateness in question does not, however, diminish the value of the theory. It was pointed out in § 60 that we can account for the observed phenomena by one solution of the field equations just as well as by another.

Now that we know that the choice of the solution is immaterial, we may simplify by some suitable restriction, — for example, by

some assumption with regard to the numerical value of the velocity of light. Two velocities, k_1 and k_2, for rays in the direction of r and at right angles to it, have to be distinguished, and for these we find, by putting $ds = 0$,

$$k_1^2 = \frac{w}{u}, \quad k_2^2 = \frac{r^2 w}{v}.$$

The special values of the potentials given in the text involve the relation

$$k_2^2 = ck_1,$$

and, conversely, we may deduce them from the general formulæ (401) if we postulate this relation, with the addition that for $r = \infty$ both k_1 and k_2 shall tend to the limit c. For then we are immediately led to the equation

$$w'^2 = \frac{c^4\alpha^2}{r^4}, \quad \text{where} \quad \alpha = \frac{1}{c}\sqrt{\frac{a}{b}}.$$

We shall take the positive value

$$w' = \frac{c^2\alpha}{r^2},$$

giving

$$w = p - \frac{c^2\alpha}{r},$$

with the integration constant p, and

$$k_2^2 = \frac{1}{a}\,[(b-p)r + c^2\alpha]^2\left(p - \frac{c^2\alpha}{r}\right).$$

This will take the value c^2 for $r = \infty$, if $p = b$. Finally, taking $a = c^4\alpha^2$, and therefore $b = c^2$,

$$w = c^2\left(1 - \frac{\alpha}{r}\right), \quad u = \frac{1}{1 - \frac{\alpha}{r}}, \quad v = r^2.$$

As we have already remarked, the angular velocity ω with which a circular orbit of radius r can be described is now determined by

$$\omega^2 = \frac{\alpha c^2}{2r^3}.$$

This shows that we were right in choosing the positive sign in the formula for w'. With the negative sign a circular orbit would have been impossible; we should then have had a repulsion instead of an attraction. Our last formula further leads to the important conclusion that in the solar system the numerical value of α/r is much

smaller than 1. For this value is seen to be equal to twice the square of the ratio between the velocity ωr in a circular orbit and the speed of light c, and this ratio is small, even immediately near the sun's surface.

For $\alpha = 0$ the values of u, v, w become 1, r^2, and c^2; this means

$$ds^2 = - d\sigma^2 + c^2dt^2,$$

corresponding to what we have called the normal values of the potentials g_{ab}.

It remains to examine the connection between the constant α and the mass of the attracting body. For this purpose we must apply the field equations not only to the external space but also to the interior of the body. If we suppose the latter to be at rest in the system of coördinates and to have a structure symmetrical around the center O, we may still use the equations (398). We shall now simplify the formulæ by assuming that u, v, w differ very little from their normal values, so that, if we put

$$u = 1 + \lambda, \quad v = r^2(1 + \mu), \quad w = c^2(1 + \nu),$$

squares and products of the quantities λ, μ, ν and their derivatives with respect to r may be neglected. This amounts to considering as very small all terms containing the gravitation constant κ; we shall therefore also neglect terms with κ^2 or with κ multiplied by λ, μ, or ν.

The expressions (399) for G_{ab} which appear on one side of the field equations

$$G_{ab} = - \kappa(T_{ab} - \tfrac{1}{2}\, g_{ab}T) \tag{402}$$

are now easily expressed in terms of λ, μ, ν and their derivatives. On the right-hand side of these formulæ the most important quantity is T_4^4, which is to represent the mass of the attractive system. Let ρ be the density. Then the mass in the element of volume corresponding to $dx_1dx_2dx_3$ will be

$$\rho x_1^2 \sin x_2\, dx_1dx_2dx_3,$$

which may be said to correspond to an energy

$$c^2\rho x_1^2 \sin x_2\, dx_1dx_2dx_3.$$

We shall therefore put

$$T_4^4 = c^2\rho x_1^2 \sin x_2. \tag{403}$$

Now, if nothing were to be neglected, we should have to take into consideration some other components of the tensor T_a^b; namely, the stresses that are developed in the body by its own gravitation,

and without which its parts cannot be in equilibrium. Since, however, these stresses are proportional to κ, and since there is already a factor κ on the right-hand side of the equations, we may confine ourselves to T_4^4.

The first thing to be done now is to deduce from (403), by means of the formulæ of §§ 58 and 59, the values of T_{ab} and T and to substitute them in (402). In all this we may give to the potentials g_{ab} their normal values, corresponding to $u = 1$, $v = r^2$, $w = c^2$. Thus:

$$g_{11} = -1, \quad g_{22} = -x_1^2, \quad g_{33} = -x_1^2 \sin^2 x_2, \quad g_{44} = c^2,$$

$$\sqrt{-g} = cx_1^2 \sin x_2.$$

Performing all these calculations, we are led to the values

$$T_{44} = c^3\rho, \quad T = c\rho$$

and to three equations (two of the four again becoming identical)

$$G_{11} = -\tfrac{1}{2} c\kappa\rho, \quad G_{22} = -\tfrac{1}{2} c\kappa\rho r^2, \quad G_{44} = -\tfrac{1}{2} c^3\kappa\rho,$$

or, when the expressions on the left-hand side are worked out,

$$-\frac{\lambda'}{r} + \frac{2\,\mu'}{r} + \mu'' + \frac{1}{2}\,\nu'' = -\frac{1}{2}\,c\kappa\rho,$$

$$-\lambda + \mu - \tfrac{1}{2}\,r\lambda' + 2\,r\mu' + \tfrac{1}{2}\,r^2\mu'' + \tfrac{1}{2}\,r\nu' = -\tfrac{1}{2}\,c\kappa\rho r^2,$$

$$c^2\left(-\frac{\nu'}{r} - \frac{1}{2}\,\nu''\right) = -\frac{1}{2}\,c^3\kappa\rho.$$

In these formulæ ρ is to be considered as a given function of r; we shall suppose it to be zero beyond a certain distance R from the center, the radius of the attracting body. The third equation will presently serve us for the determination of ν. After we have solved it, the relation

$$\lambda = \mu + r\mu' + r\nu' \tag{404}$$

will be found sufficient for satisfying the first and the second equation. That the problem is again indeterminate will not be astonishing after what has been said.

Now, confining ourselves to the third equation, we find in the first place, after multiplication by $-\dfrac{2\,r^2}{c^2}$,

$$\frac{d}{dr}\,(r^2\nu') = c\kappa\rho r^2,$$

$$r^2\nu' = c\kappa\int_0^r \rho r^2 dr,$$

where the integration constant has been determined by the condition that $r^2\nu'$ must vanish at the center. Moreover, taking into account that ν must be zero at infinite distance,

$$\nu = c\kappa\int_{\infty}^{r}\frac{dr}{r^2}\int_0^r \rho r^2 dr.$$

or, after an integration by parts,

$$\nu = -\frac{c\kappa}{r}\int_0^r \rho r^2 dr - c\kappa\int_r^{\infty}\rho r\, dr.$$

At a point outside the attracting body the second integral vanishes and the first has the value

$$\int_0^R \rho r^2 dr = \frac{m}{4\,\pi}$$

if m is the total mass. Thus

$$\nu = -\frac{c\kappa m}{4\,\pi r}, \quad w = c^2\left(1 - \frac{c\kappa m}{4\,\pi r}\right),$$

so that the coefficient α formerly introduced is connected with the mass m by the relation

$$\alpha = \frac{c\kappa m}{4\,\pi}.$$

We shall conclude by remarking that among the different solutions of the field equations considered in this Note there is one in which, in the expression for ds^2, the squares of the components $d_1\sigma$ and $d_2\sigma$ of the line-element in R_3 are multiplied by the same coefficient, so that the fundamental formula takes the form

$$ds^2 = -\alpha^2 d\sigma^2 + \beta^2 dx_4^2,$$

or, if x, y, z are rectangular coördinates,

$$ds^2 = -\alpha^2(dx^2 + dy^2 + dz^2) + \beta^2 dt^2. \tag{405}$$

For this, all that is necessary is that

$$v = r^2 u,$$

or, in the later formulæ, $\lambda = \mu.$

If the first relation is combined with (401), a simple integration will enable us to determine w, and thereby u and v as functions of r. On the other hand, if $\lambda = \mu$, equation (404) can be satisfied by

$$\lambda = \mu = -\nu,$$

where ν is to be given the value that we have already found.

Note 24, § 61

The problem of the influence of gravitation on the frequency of the light emitted by an atom may be treated in connection with Bohr's theory, in which the frequency is made to depend on a change of energy.

Let us use a system of reference x, y, z, t, in which x, y, z are rectangular coördinates extending all through the solar system. We shall consider solely the gravitational field due to the sun, and we shall admit for it equation (405), mentioned at the end of Note 23. Let α, β be the values of the constants at the sun's surface, and α_0, β_0 the values near the earth.

We shall begin by fixing our attention on an atom in our laboratory and on one of its stationary states of motion. The coördinates of each electron will be definite functions of the time

$$x = \phi(t), \quad y = \psi(t), \quad z = \chi(t),$$

and the atom will have a certain energy E_0.

Our next step is to describe this *same* state of things in terms of new variables defined by

$$x' = \frac{\alpha_0}{\alpha}\,x, \quad y' = \frac{\alpha_0}{\alpha}\,y, \quad z' = \frac{\alpha_0}{\alpha}\,z, \quad t' = \frac{\beta_0}{\beta}\,t, \tag{406}$$

so that, always near the earth, the fundamental formula becomes

$$ds'^2 = -\,\alpha^2(dx'^2 + dy'^2 + dz'^2) + \beta^2 dt'^2. \tag{407}$$

In the new system of reference the equations for the stationary motions are

$$x' = \frac{\alpha_0}{\alpha}\,\phi\left(\frac{\beta}{\beta_0}\,t'\right), \quad y' = \frac{\alpha_0}{\alpha}\,\psi\left(\frac{\beta}{\beta_0}\,t'\right), \quad z' = \frac{\alpha_0}{\alpha}\,\chi\left(\frac{\beta}{\beta_0}\,t'\right), \tag{408}$$

and the energy will have a value E' which we can deduce from E_0 by means of the formulæ of transformation.

In the case of a single moving particle the energy is the fourth component of a covariant vector q_a, so that, with the above relations between x, y, z, t and x', y', z', t',

$$q_4' = \frac{\beta}{\beta_0}\,q_4.$$

If the theory is to be self-consistent, the same formula must hold for the energy of the atom. This may be verified by considering the integral

$$\int T_4^4\,dx\,dy\,dz$$

and using the transformation formula for T_a^b.

Thus
$$E' = \frac{\beta}{\beta_0}\,E_0. \tag{409}$$

We shall now use the principle of relativity in the following form. If in two systems of reference x, y, z, t and x', y', z', t' the gravitation potentials have the same values, there will be for any given phenomenon a corresponding one, equally possible, that is described by equations of exactly the same form in x', y', z', t' as the first in x, y, z, t, and conversely.

Now we see from (407) that near the earth the potentials have in x', y', z', t' the values α, β which on the sun they have in x, y, z, t. Thus, since on the earth we have recognized as possible motions those specified by (408), whose energy is given by (409), we conclude that an atom on the sun can be in a state of motion determined by

$$x = \frac{\alpha_0}{\alpha}\,\phi\left(\frac{\beta}{\beta_0}\,t\right), \quad y = \frac{\alpha_0}{\alpha}\,\psi\left(\frac{\beta}{\beta_0}\,t\right), \quad z = \frac{\alpha_0}{\alpha}\,\chi\left(\frac{\beta}{\beta_0}\,t\right)$$

and characterized by the energy

$$E = \frac{\beta}{\beta_0}\,E_0.$$

So, if both particles are referred to one and the same system x, y, z, t, the ratio between the energies of two atoms, one on the sun and the other on the earth, taken in corresponding stationary states, is found to be β/β_0. The same ratio must exist between corresponding differences in energy, and therefore, if Planck's constant h is given the same value in the two cases, between the emitted frequency as determined by Bohr's rule. It will easily be seen that this agrees with what was said in the text.

That h must be considered as equal in two places with different gravitation potentials is due to the fact that h is the product of an energy and a time. Attending to the way in which the numerical values of these quantities are altered when we pass from one system of reference to the other, we find that the value of h does not change.

In all this there is something that looks paradoxical at first sight. The change from x, y, z, t to x', y', z', t' defined by (406) is simply what we should have had if we had changed our units of length and time, so that the new units are equal to α/α_0 and β/β_0 times the corresponding old ones. By this operation, as (409) shows, the numerical value of the energy changes in the ratio of 1 to β/β_0. How can this be reconciled with the fact that the dimensions of energy are ML/T^2? The answer is that if the transformation formulæ which we have used are to hold for a transformation like (406), amounting to a change of units, we must change the unit of mass in the ratio of

1 to p/q when the unit of length and that of time are altered in the ratios of 1 to q and 1 to p respectively. In other words, we must attribute to the mass the dimensions T/L. By this the dimensions of an energy become $1/T$, and h becomes a pure number without dimensions.

Note 25, § 63

It has been remarked in the text that the values of the potentials are different for two systems of coördinates in one of which the earth rotates with a constant velocity, while the other is rigidly connected with it. In the first system the potentials may have the normal values; in the second they are such as to give rise to phenomena like those of Foucault's pendulum.

The question has also been raised, What is it with respect to which the earth moves? By the old theories we could imagine that it rotates relatively to the ether of space, and could suppose that the coefficients g_{ab} have their normal values when the axes of coördinates are at rest in the ether, and different ones when they are rotating in it.

Einstein has, however, succeeded in replacing, in this line of thought, the ether by ordinary matter; namely, by the heavenly bodies existing in the universe. He showed that these bodies can be made accountable for the values of the gravitation potentials even when these are as simple as the normal ones, so that we must necessarily have these values when the axes of coördinates are at rest relatively to the bodies in question. In all this we must not think of some new action exerted by matter but only of the influence which it has according to the field equations.

For his purpose Einstein had to make two changes in the original theory. The first was the modification of the field equations that has already been mentioned. The new form is

$$G_{ab} - \lambda g_{ab} = -\kappa(T_{ab} - \tfrac{1}{2}\, g_{ab}T), \tag{410}$$

where λ is a certain positive constant. The second change is that, for a reason which must be left unexplained here, the three-dimensional space R_3 (namely, the space in the ordinary sense of the word) is supposed to be of finite extent, though it has no boundary. This new kind of space is the three-dimensional counterpart of a spherical surface. Like this it has the same properties in all its parts; and just as the spherical surface can be defined in a three-dimensional extension by the equation

$$x^2 + y^2 + z^2 = a^2,$$

so Einstein's space R_3 can be defined in a four-dimensional extension q, x, y, z (not to be confounded with the extension x, y, z, t) by the equation

$$q^2 + x^2 + y^2 + z^2 = a^2. \tag{411}$$

The constant a may be called its radius.

The position of a point on the spherical surface may be determined by means of two angles ϕ and ψ, such that

$$z = a \cos \phi, \quad x = a \sin \phi \cos \psi, \quad y = a \sin \phi \sin \psi;$$

and if the values of dx, dy, dz derived from these formulæ, for a constant a, are substituted in

$$d\sigma^2 = dx^2 + dy^2 + dz^2,$$

we find, for the square of the line-element,

$$d\sigma^2 = a^2(d\phi^2 + \sin^2 \phi \, d\psi^2).$$

Similarly, any point of Einstein's R_3 can be represented by three angles ϕ, ψ, χ such that

$$\left.\begin{aligned} q &= a \cos \phi \\ x &= a \sin \phi \cos \psi \\ y &= a \sin \phi \sin \psi \cos \chi \\ z &= a \sin \phi \sin \psi \sin \chi \end{aligned}\right\}, \tag{412}$$

equation (411) being identically satisfied by these values; and if the values of dq, dx, dy, dz are substituted in

$$d\sigma^2 = dq^2 + dx^2 + dy^2 + dz^2,$$

we find, for the square of the distance of neighboring points of R_3,

$$d\sigma^2 = a^2(d\phi^2 + \sin^2 \phi \, d\psi^2 + \sin^2 \phi \sin^2 \psi \, d\chi^2),$$

or, if we take ϕ, ψ, χ as coördinates,

$$x_1 = \phi, \quad x_2 = \psi, \quad x_3 = \chi,$$

$$d\sigma^2 = a^2(dx_1^2 + \sin^2 x_1 dx_2^2 + \sin^2 x_1 \sin^2 x_2 dx_3^2). \tag{413}$$

Having gone so far, we need no longer think of the four-dimensional space q, x, y, z, the introduction of which could even have been entirely avoided. It suffices to state that Einstein's space is a three-dimensional extension the geometry of which is made to depend on this formula for its line-element. The symbol a denotes a certain length that may be expressed in centimeters, and we shall travel through the total extent of this R_3 when x_1 and x_2 are made to change from 0 to π and x_3 from 0 to 2π.

Much of what has been said in Note 23 may, with appropriate changes, be applied to the present space R_3. From any of its points we can draw three line-elements in the directions of the coördinates (two coördinates remaining constant along each element) and characterized by dx_1, dx_2, dx_3. According to (413) their lengths are

$$a\,dx_1, \quad a \sin x_1 dx_2, \quad a \sin x_1 \sin x_2 dx_3,$$

and they are at right angles to each other. It is therefore natural to take the product of these three expressions

$$a^3 \sin^2 x_1 \sin x_2 dx_1 dx_2 dx_3$$

as the measure of the magnitude of the element of volume which they determine. Integrating this between the limits given just now, we find, for the total magnitude of R_3,

$$2\,\pi^2 a^3.$$

We shall replace the matter concentrated in the stars by a uniform distribution throughout R_3 with density ρ. Then the mass contained in an element becomes

$$\rho a^3 \sin^2 x_1 \sin x_2 dx_1 dx_2 dx_3,$$

and we therefore put

$$T_4^4 = c^2 \rho a^3 \sin^2 x_1 \sin x_2.$$

We shall suppose the other components T_a^b to be all zero, the matter being at rest and without internal stresses, and we proceed to show that the field equations can be satisfied if we assume

$$ds^2 = -u d\sigma^2 + w dt^2$$

with suitably chosen constant values of u and w.

Substituting the value of $d\sigma^2$, we see that

$$g_{11} = -a^2 u, \;\; g_{22} = -a^2 u \sin^2 x_1, \;\; g_{33} = -a^2 u \sin^2 x_1 \sin^2 x_2, \;\; g_{44} = w,$$

$$\sqrt{-g} = a^3 \sqrt{u^3 w} \cdot \sin^2 x_1 \sin x_2,$$

giving

$$G_{11} = -2, \;\; G_{22} = -2 \sin^2 x_1, \;\; G_{33} = -2 \sin^2 x_1 \sin^2 x_2, \;\; G_{44} = 0,$$

$$T_{44} = c^2 \rho \sqrt{\frac{w}{u^3}}, \quad T = \frac{c^2 \rho}{\sqrt{u^3 w}}.$$

Of the four field equations, with $G_{11} \cdots G_{44}$, each of the first three leads to the formula

$$\frac{2}{a^2 u} - \lambda = \frac{1}{2} \kappa T,$$

and the fourth to

$$\lambda = \tfrac{1}{2} \kappa T,$$

so that u and w, and thereby the gravitation potentials, are really determined when a, ρ, κ, and λ are given. If the potentials are to have the normal values $u = 1$, $w = c^2$, we must take

$$\lambda = \frac{1}{a^2} \quad \text{and} \quad \rho = \frac{2}{c\kappa a^2}.$$

In what follows we shall suppose these conditions to be satisfied, so that the field is represented by

$$ds^2 = -a^2(dx_1^2 + \sin^2 x_1 d\psi^2 + \sin^2 x_1 \sin^2 \psi\, d\chi^2) + c^2dt^2. \quad (414)$$

In the space R_3 there is a definite point O where x_1 or ϕ is zero (compare equations 412), and from which this angle x_1 is measured. Now let us consider around this point a part R_3' of R_3 whose dimensions are, for example, a small number of times those of the solar system. Then they will of course be extremely small with respect to the radius a of the world, and the angle x_1 will be so small that $\sin x_1$ may be replaced by x_1 to a high degree of approximation. By this the above formula becomes, if we write r for ax_1,

$$ds^2 = -dr^2 - r^2(d\psi^2 + \sin^2 \psi\, d\chi^2) + c^2dt^2. \quad (415)$$

Obviously the quantity r is the distance from O, and so we find for the smaller space R_3' the simplest form which ds^2, and the gravitation potentials included in it, can take when polar coördinates are used. To judge by this form we should say that the space R_3' is that of Euclidean geometry, and that it is the seat of the simplest gravitational field that we can imagine, or rather of a field of no gravitation. Now conceive the earth to be placed at the center O of this field. In doing so we shall not trouble ourselves with the gravitation due to the earth or even with that of the sun; as has been explained, the actions of these bodies would produce only very slight changes in the potentials, and we are now interested only in their main values.

As these are the normal ones, the system of coördinates to which (415) refers is one in which the earth is rotating, with a certain angular velocity ω, say about the axis from which the angle ψ is reckoned. If the angular velocity is ω, we can pass to axes fixed to the earth by simply introducing instead of χ a new coördinate

$$\chi' = \chi - \omega t,$$

leaving r and ψ unchanged. The formula for the line-element now becomes

$$\begin{aligned} ds'^2 = -dr^2 - r^2(d\psi^2 + \sin^2 \psi\, d\chi'^2) - 2\, r^2\omega \sin^2 \psi\, d\chi' dt \\ + (c^2 - r^2\omega^2 \sin^2 \psi)dt^2, \end{aligned} \quad (416)$$

showing us the gravitation potentials, with a g_{34} among them, such as they are relatively to axes that are fixed to the earth. If we use these, the new potentials will serve us for the explanation, for example, of Foucault's pendulum experiment. Now, just as field (415) in R_3' is part of the complete field (414), so also field (416) in R_3' will be part of a field extending all through R_3, the fundamental formula for which is obtained when, in (414), χ is replaced by $\chi' + \omega t$, so that the potentials become

$$\left.\begin{array}{l} g'_{11} = -a^2, \quad g'_{22} = -a^2 \sin^2 x_1, \quad g'_{33} = -a^2 \sin^2 x_1 \sin^2 x_2 \\ g'_{34} = -a^2\omega \sin^2 x_1 \sin^2 x_2, \quad g'_{44} = c^2 - a^2\omega^2 \sin^2 x_1 \sin^2 x_2 \end{array}\right\}, \qquad (417)$$

different from the former values

$$g_{11} = -a^2, \quad g_{22} = -a^2 \sin^2 x_1, \quad g_{33} = -a^2 \sin^2 x_1 \sin^2 x_2,$$
$$g_{34} = 0, \quad g_{44} = c^2.$$

Like these the new values g'_{ab} can be deduced from the field equations, which have the form (410) in the system ϕ, ψ, χ' as well as in $\phi, \psi\ \chi$; namely,

$$G'_{ab} - \lambda g'_{ab} = -\kappa(T'_{ab} - \tfrac{1}{2}\, g'_{ab}T'). \qquad (418)$$

Only, if we want to do so, we must not forget that in the system ϕ, ψ, χ' the stars are moving. Indeed, for each star χ was supposed to be constant, but then the new coördinate χ', or x_3' decreases at the rate ω. The effect will be that in the system of reference x_1, x_2, x_3' the energy $T_4'^4$ is not the only component of the stress-energy tensor with which we are concerned; it is clear that there will now be a momentum and a flow of energy.

Now if the right values of T'_{ab} are substituted in (418), we shall find the potentials (417) as a solution of the equations. We shall in this way be led to the expression (416), with which we have to reckon in the space R_3'. We may say, therefore, that the phenomena by which the rotation of the earth is usually proved are due to the fact that in a system of reference in which the earth is at rest the stars are moving around it.

NAME INDEX

SUBJECT INDEX